

N° d'ordre : 2015-ISAL-0040

THESE

Sonie de champs acoustiques stationnaires en situation d'écoute dichotique

présentée devant

L'Institut National des Sciences Appliquées de Lyon

pour obtenir

Le grade de docteur

Ecole doctorale

Mécanique, Energétique, Génie civil, Acoustique (MEGA)

Spécialité : Acoustique

par

Michaël VANNIER

Soutenue publiquement le 11 mai 2015

Jury :

Wolfgang ELLERMEIER	Professeur des universités, TU, Darmstadt	Rapporteur
Nicolas GRIMAULT	Chargé de recherche, CNRS, Lyon	Rapporteur
Vincent KOEHL	Maître de conférences, UBO, Brest	Examineur
Sabine MEUNIER	Chargée de recherche, CNRS, Marseille	Examineur
Etienne PARIZET	Professeur des universités, INSA, Lyon	Directeur de thèse
Daniel PRESSNITZER	Directeur de recherche, CNRS, Paris	Examineur

Michaël VANNIER

Sonie de champs acoustiques stationnaires en situation d'écoute dichotique

Thèse préparée au Laboratoire Vibrations Acoustique de l'INSA de Lyon,
Dans le cadre du LabEx CeLyA (Centre Lyonnais d'Acoustique, ANR-10-LABX-60),
Soutenue par l'ANR (projet "Loudnat", convention ANR-11-BS09-016-01).

INSA Direction de la Recherche - Ecoles Doctorales – Quinquennal 2011-2015

SIGLE	ECOLE DOCTORALE	NOM ET COORDONNEES DU RESPONSABLE
CHIMIE	CHIMIE DE LYON http://www.edchimie-lyon.fr Sec : Renée EL MELHEM Bat Blaise Pascal 3 ^e etage 04 72 43 80 46 Insa : R. GOURDON secretariat@edchimie-lyon.fr	M. Jean Marc LANCELIN Université de Lyon – Collège Doctoral Bât ESCPE 43 bd du 11 novembre 1918 69622 VILLEURBANNE Cedex Tél : 04.72.43 13 95 directeur@edchimie-lyon.fr
E.E.A.	ELECTRONIQUE, ELECTROTECHNIQUE, AUTOMATIQUE http://edeea.ec-lyon.fr Sec : M.C. HAVGOUDOUKIAN Ecole-doctorale.eea@ec-lyon.fr	M. Gérard SCORLETTI Ecole Centrale de Lyon 36 avenue Guy de Collongue 69134 ECULLY Tél : 04.72.18 60.97 Fax : 04 78 43 37 17 Gerard.scorletti@ec-lyon.fr
E2M2	EVOLUTION, ECOSYSTEME, MICROBIOLOGIE, MODELISATION http://e2m2.universite-lyon.fr Sec : Safia AIT CHALAL Bat Atrium- UCB Lyon 1 04.72.44.83.62 Insa : S. REVERCHON Safia.ait-chalal@univ-lyon1.fr	M. Fabrice CORDEY Laboratoire de Géologie de Lyon Université Claude Bernard Lyon 1 Bât Géode – Bureau 225 43 bd du 11 novembre 1918 69622 VILLEURBANNE Cédex Tél : 04.72.44.83.74 Sylvie.reverchon-pescheux@insa-lyon.fr fabrice.cordev@univ-lyon1.fr
EDISS	INTERDISCIPLINAIRE SCIENCES-SANTE http://www.ediss-lyon.fr Sec : Safia AIT CHALAL Bat Atrium – UCB Lyon 1 04 72 44 83 62 Insa : Safia.ait-chalal@univ-lyon1.fr	Mme Emmanuelle CANET-SOULAS INSERM U1060, CarMeN lab, Univ. Lyon 1 Bâtiment IMBL 11 avenue Jean Capelle INSA de Lyon 696621 Villeurbanne Tél : 04.72.11.90.13 Emmanuelle.canet@univ-lyon1.fr
INFOMATHS	INFORMATIQUE ET MATHEMATIQUES http://infomaths.univ-lyon1.fr Sec : Renée EL MELHEM Bat Blaise Pascal 3 ^e etage infomaths@univ-lyon1.fr	Mme Sylvie CALABRETTO LIRIS – INSA de Lyon Bat Blaise Pascal 7 avenue Jean Capelle 69622 VILLEURBANNE Cedex Tél : 04.72. 43. 80. 46 Fax 04 72 43 16 87 Sylvie.calabretto@insa-lyon.fr
Matériaux	MATERIAUX DE LYON http://ed34.universite-lyon.fr Sec : M. LABOUNE PM : 71.70 –Fax : 87.12 Bat. Saint Exupéry Ed.materiaux@insa-lyon.fr	M. Jean-Yves BUFFIERE INSA de Lyon MATEIS Bâtiment Saint Exupéry 7 avenue Jean Capelle 69621 VILLEURBANNE Cedex Tél : 04.72.43 71.70 Fax 04 72 43 85 28 Ed.materiaux@insa-lyon.fr
MEGA	MECANIQUE, ENERGETIQUE, GENIE CIVIL, ACOUSTIQUE http://mega.universite-lyon.fr Sec : M. LABOUNE PM : 71.70 –Fax : 87.12 Bat. Saint Exupéry mega@insa-lyon.fr	M. Philippe BOISSE INSA de Lyon Laboratoire LAMCOS Bâtiment Jacquard 25 bis avenue Jean Capelle 69621 VILLEURBANNE Cedex Tél : 04.72. 43.71.70 Fax : 04 72 43 72 37 Philippe.boisse@insa-lyon.fr
ScSo	ScSo* http://recherche.univ-lyon2.fr/scso/ Sec : Viviane POLSINELLI Brigitte DUBOIS Insa : J.Y. TOUSSAINT viviane.polsinelli@univ-lyon2.fr	Mme Isabelle VON BUELTZINGLOEWEN Université Lyon 2 86 rue Pasteur 69365 LYON Cedex 07 Tél : 04.78.77.23.86 Fax : 04.37.28.04.48 isavonb@dbmail.com

*ScSo : Histoire, Géographie, Aménagement, Urbanisme, Archéologie, Science politique, Sociologie, Anthropologie

Remerciements

Je commencerai par remercier Etienne Parizet, pour la confiance qu'il m'a accordée tout au long de ce travail, son recul nécessaire, sa disponibilité et pour les longues heures consacrées à l'écoute de bruits spatialisés.

J'adresse également mes remerciements à l'ensemble des membres du jury. Ainsi je remercie Nicolas Grimault et Wolfgang Ellermeier pour m'avoir fait l'honneur de rapporter ce travail, Sabine Meunier et Vincent Koehl pour avoir accepté de faire partie du jury et Daniel Pressnitzer pour l'avoir présidé.

Je suis d'autre part très heureux d'avoir pu participer au projet ANR LoudNat et tiens à remercier l'ensemble des participants et initiateurs de ce projet, ayant permis de riches échanges scientifiques (merci notamment pour le workshop !), dans un cadre de travail toujours convivial et chaleureux. Merci donc à l'équipe du LMA : Sabine Meunier, Jacques Chatron, Guy Rabau et Sophie Savel, ainsi qu'à l'équipe de l'IRCAM : Patrick Susini et Emmanuel Ponsot. Je tiens à remercier plus particulièrement Sabine et Patrick pour m'avoir initié et transmis leur intérêt pour la recherche et pour m'avoir toujours soutenu dans cette voie. Je dois également beaucoup à Emmanuel Ponsot, à ses nombreux conseils avisés et à nos longues discussions sur la sonie.

Je tiens également à remercier très sincèrement Wolfgang Ellermeier pour son accueil chaleureux dans son laboratoire, à l'université de Darmstadt, ainsi que pour les échanges très constructifs avec Josef Schlittenlacher qui m'ont beaucoup apporté pour la thèse. Je remercie également Jana Birkenbusch et l'ensemble des étudiants et personnels de l'institut de psychologie pour leur accueil, les échanges et l'esprit d'équipe dans la "TU meet & moove" !

Comme dans l'ensemble des études en psychophysique, cette thèse n'aurait pas été possible sans l'engagement volontaire de chacun des auditeurs. La plupart d'entre eux a été remarquablement assidue pendant 3 ans, ce qui m'a permis d'avoir un suivi tout au long du travail de thèse. Je remercie donc chacun des auditeurs ayant prêté leur temps et leurs oreilles pour mes expériences, à Lyon comme à Darmstadt.

Un grand merci à l'ensemble des permanents et des doctorants du LVA pour leurs conseils, leur soutien amical et leur convivialité, ayant contribué à m'offrir un cadre de travail des plus agréables. Merci notamment à Laurent Brocolini ainsi qu'à mes collègues de bureau (et amis) Thibault Lafont, Thibaut le Magueresse et Jérémy Moriot pour leur soutien, leur aide et leur bonne humeur au quotidien.

Pour finir, je souhaiterais remercier ma famille, notamment mes parents pour leur soutien depuis toujours. Une dédicace spéciale à grand-maman, qui de son point de vue de scientifique, a contribué de manière importante à la finalisation du manuscrit. Enfin, un immense merci à ma loupote pour son soutien chaque jour, inestimable.

Au souvenir de mon grand-père,

Table des matières

Introduction générale	1
1 Contexte de l'étude	5
1.1 Sonie des sons stationnaires : principes fondamentaux	5
1.1.1 Définitions	5
1.1.2 Les principaux facteurs qui influent sur la sonie	5
1.1.2.1 L'intensité	6
1.1.2.2 La fréquence	8
1.1.2.3 La largeur de bande (intégration spectrale)	9
1.1.2.4 La durée (intégration temporelle)	10
1.2 Sonie binaurale	11
1.2.1 Sommation de sonie binaurale	11
1.2.1.1 Comment peut-on quantifier la sommation de sonie binaurale ?	11
1.2.1.2 Facteurs influençant la sommation de sonie binaurale	15
1.3 Sonie binaurale d'un champ acoustique	17
1.3.1 Sonie binaurale d'une source en champ libre	17
1.3.2 Sonie binaurale de plusieurs sources en champ libre	20
1.3.3 Sonie binaurale en-dehors du champ libre	21
1.3.4 Lorsque la sonie ne dépend plus uniquement des signaux reçus aux oreilles	23
1.3.4.1 Constance perceptuelle de la sonie (<i>Loudness constancy</i>)	23
1.3.4.2 Interactions multisensorielles	24
1.3.4.3 Sonie et Analyse de Scène Auditive (ASA)	25
Conclusion	26
1.4 Modèles de sonie binaurale	26
1.4.1 Présentation des différents modèles	26
1.4.1.1 Calcul de sonie en présentation diotique	26
Étape 1 :	27
Étapes 2 et 3 :	28
Étapes 4 et 5 :	28
1.4.1.2 Modélisation de la sommation de sonie binaurale	29
a) Sommation de sonie parfaite :	29
b) Modèle de Moore et Glasberg (2007) :	29
c) Modèle de Sivonen et Ellermeier (2008) :	32
1.4.2 Validité et limitations	33
1.4.2.1 Domaines de validité	33
1.4.2.2 Influence du choix des HRTFs	35

I	Sonie directionnelle et variabilité interindividuelle	41
2	Mesures de sonie directionnelle pour une source de bruit à bande étroite	43
2.1	Méthode	43
2.1.1	Auditeurs	43
2.1.2	Stimuli	43
2.1.3	Procédure	44
2.1.4	Dispositif expérimental	45
2.1.5	Mesure des niveaux SPL	46
2.2	Résultats	46
2.2.1	Résultats individuels	46
2.2.2	Groupes d'auditeurs	48
2.2.3	DLS médianes et niveaux SPL " aux tympanes "	51
2.3	Modélisation	54
2.3.1	Modélisation des données expérimentales	54
2.3.2	Comparaison de différents gains de sommation binaurale	58
2.4	Discussion	60
2.5	Conclusion	62
3	Influence de la variabilité des fonctions de sonie monaurales individuelles	65
3.1	Mesure des fonctions de sonie par estimation de grandeur (<i>ME</i>)	65
3.1.1	Méthode	65
3.1.1.1	Auditeurs	65
3.1.1.2	Stimuli	65
3.1.1.3	Procédure	66
3.1.2	Résultats	67
3.1.2.1	Fonctions de sonie, méthode de Cross	67
3.1.2.2	Comparaison des sessions de mesure 1 et 2	69
3.1.2.3	Résultats sur l'ensemble des deux sessions	71
3.2	Mesure des fonctions de sonie par production de rapport (<i>RP</i>)	73
3.2.1	Méthode	74
3.2.1.1	Stimuli	74
3.2.1.2	Auditeurs	74
3.2.1.3	Procédure	74
3.2.2	Résultats	75
3.3	Discussion sur les exposants, comparaison des deux méthodes	78
3.4	Exposants individuels et DLS	81
3.5	Conclusion	85
II	Sonie binaurale dans un champ acoustique stationnaire complexe	87
4	Sonie d'un champ acoustique multi-tonal, composé d'une source bitonale ou de deux sources tonales distribuées	89

4.1	Introduction	89
4.2	Méthode générale	90
4.2.1	Auditeurs	90
4.2.2	Dispositif expérimental	90
4.2.3	Stimuli	90
4.2.4	Procédure	92
4.3	Sonie directionnelle d'une source bitonale	93
4.3.1	Résultats	93
4.3.2	Discussion	94
4.4	Sonie de deux sources tonales simultanées, distribuées dans le plan horizontal	95
4.4.1	Résultats	95
4.4.2	Dissymétrie par rapport au plan sagittal (G/D)	96
4.4.3	Répétition de l'expérience avec une nouvelle égalisation du casque	98
4.4.3.1	Méthode	98
4.4.3.2	Résultats	99
4.4.4	Distribution des sources et ségrégation	100
4.4.5	Conclusion	103
4.5	De la sonie "relative" à la sonie "absolue" d'un champ acoustique multi-tonal	104
4.5.1	Comparaison de sonie entre une source bitonale (f_1+f_2) et une source tonale (f_1 ou f_2)	104
4.5.1.1	Résultats	104
4.5.1.2	Discussion	105
4.5.2	Comparaison de sonie entre une source bitonale (f_1+f_2) et une source tonale (f_3)	108
4.5.2.1	Méthode	108
4.5.2.2	Résultats	109
4.5.2.3	Discussion	110
4.6	Modélisation	111
4.6.1	Prédiction de la sonie "relative" d'un champ acoustique bitonal	112
4.6.1.1	Optimisation du gain binaural pour le modèle de Sivonen et Ellermeier (2008)	112
4.6.1.2	Comparaison des différents modèles	113
4.6.2	Prédiction de la sonie "absolue" de la source bitonale de référence (A)	116
4.7	Conclusion	119
5	Prédiction de la sonie d'un champ acoustique complexe, composé d'une ou plusieurs sources réelles.	121
5.1	Méthode	121
5.1.1	Auditeurs	121
5.1.2	Dispositif expérimental	121
5.1.3	Stimuli	122
5.1.4	Procédure	124
5.2	Résultats et discussion	125
5.3	Validité des modèles	130

5.4 Conclusion	134
Conclusion générale	135
A Audiogrammes	141
A.1 Laboratoire Vibrations Acoustique, INSA-Lyon	141
A.2 Institut für psychologie, TU-Darmstadt	143
B Synthèse binaurale générique	145
B.1 Principe général	145
B.2 Mesure des réponses impulsionnelles binaurales (RIB)	147
B.2.1 Dispositif expérimental	147
B.2.2 Mesures	147
B.3 Égalisation des transducteurs	149
B.3.1 Fonctions de transfert du casque (FTCs)	149
B.3.2 Fonctions de transfert des haut-parleurs (FTHPs)	150
B.4 Construction des filtres pour la synthèse	151
B.5 Validation de la synthèse	152
B.5.1 Enregistrement en champ réel	152
B.5.2 Enregistrement à la restitution (post-synthèse)	152
B.5.3 Comparaison des enregistrements	152
C Calibration du casque	157
C.1 Calibration "aux tympans", en équivalent champ diffus	157
C.1.1 Mesure des niveaux de pression équivalents "aux tympans"	157
C.1.2 Calibration du casque	159
D Dissymétrie G/D dans les fonctions de transfert du casque, mesure et validation de nouvelles FTCs	161
D.1 Contrôle de la symétrie G/D dans l'égalisation de la réponse du casque	161
D.1.1 Méthode	161
D.1.2 Résultats	162
D.2 Mesure et validation de nouvelles fonctions de transfert	164
E Analyse spectrale des sources sonores réelles	167
Références	175

Table des abréviations

<i>AME</i>	Absolute Magnitude Estimation / Estimation de grandeur sans référence
<i>AMP</i>	Absolute Magnitude Production / Production de grandeur sans référence
ANOVA	Analysis Of VAriance / Analyse de la variance
ASA	Auditory Scene Analysis / Analyse de Scène Auditive
BELRH	Binaural Equal Loudness Ratio Hypothesis / Hypothèse d'invariance du rapport de sonie binaural-à-monaural
CMM	Cross-Modality Matching / Appariement inter-modalité
D	Droite
DLS	Directional Loudness Sensitivity / Sensibilité directionnelle en sonie
EQM	Écart Quadratique Moyen
ERB	Equivalent rectangular Bandwidth / Bande rectangulaire équivalente
FFT	Fast Fourier Transform / Transformée de Fourier rapide
FTC(I)	Fonction de Transfert du Casque (Inverse)
FTHP	Fonction de Transfert du Haut-Parleur
G	Gauche
HP	Haut-parleur
HRIR	Head-Related Impulse response / Réponse impulsionnelle relative à la tête
HRTF	Head-Related Transfer Function / Fonction de transfert relative à la tête
ILD	Interaural Loudness Difference / Différence interaurale de niveau
ISI	InterStimulus Interval / Interval Inter-Stimuli
ITD	Interaural Time Difference / Différence interaurale de temps
JND	Just Noticeable Difference / Différence juste audible
K-S	Test de Kolmogorov-Smirnov
<i>ME</i>	Magnitude Estimation / Estimation de grandeur
<i>MP</i>	Magnitude Production / Production de grandeur
PSE	Point of Subjective Equality / Point d'égalité subjective
RE	Ratio Estimation / Estimation de rapport
RIB	Réponse Impulsionnelle Binaurale
<i>RP</i>	Ratio Production / Production de rapport
TA	Tête Artificielle
TDS	Test Des Signes

Notations

TR_{60}	Temps de réverbération
Φ	Azimut
Θ	Elévation
ρ	Coefficient de corrélation de Pearson

Introduction générale

La sonie, définie comme la sensation de force sonore, est un attribut perceptif intuitif dont certains aspects peuvent être facilement appréhendés par l'individu non entraîné, notamment dans un contexte quotidien d'environnement sonore souvent trop fort, pouvant jouer sur les facultés cognitives et/ou l'état physique et mental général de l'individu, voire provoquer des pertes auditives dans le cas d'expositions longues. De nombreuses applications trouvent ainsi leur source dans la gestion des sons forts désirables (médias, musique live, alarmes...) ou indésirables (chantiers, trafic routier,...), l'évaluation de la toxicité ou la réhabilitation des malentendants. La sonie est également fortement impliquée dans l'évaluation de la qualité sonore, du désagrément ou de la gêne (e.g. Berglund *et al.* (1990)).

Selon l'approche psychophysique, l'étude de la sonie consiste à comprendre l'influence des différents paramètres physiques ou autres facteurs contextuels, sensoriels et cognitifs sur la sonie, à travers la mise en place d'expériences psychoacoustiques consistant à recueillir le jugement d'auditeurs en réponse à des stimuli. La modélisation des processus ainsi mis en évidence vise ensuite à pouvoir prédire la sonie (subjective) directement à partir d'une mesure (objective) du champ acoustique. Des modèles existent et permettent de prédire la sonie des sons stationnaires (présentant des propriétés statistiques stables au cours du temps) dans des situations d'écoute simples (ISO-532B (1975), DIN-45631 (1990) ANSI-S3.4 (2007)). L'écoute doit être monaurale (une seule oreille) ou diotique (même son aux deux oreilles), correspondant à une source en incidence frontale en champ libre, ou en champ diffus. En revanche, ces modèles échouent pour prédire la sonie lorsque des différences interaurales importantes interviennent (e.g. Sivonen et Ellermeier (2006)). D'autres modèles ont été proposés pour prédire la sonie au cours du temps pour des sons non-stationnaires (Zwicker et Fastl (1999), Glasberg et Moore (2002), Chalupper et Fastl (2002)). Mais les indicateurs proposés pour estimer la sonie globale (N_4 , N_5 , $SLTL_{max}$,...) échouent notamment pour des sons longs (plusieurs secondes) dont la distribution temporelle des niveaux n'est pas aléatoire (e.g. Meunier et Marchioni (2002), Ponsot *et al.* (2015)).

Or, dans un environnement naturel, le champ acoustique est complexe (plusieurs sources, différentes positions spatiales, acoustique du lieu,...) et varie au cours du temps (sources non-stationnaires avec des profils temporels souvent longs et asymétriques, évolution du nombre de sources,...). De plus, l'écoute étant binaurale, le filtrage acoustique opéré par la tête, le buste et les pavillons de l'auditeur (dépendant de la direction) induit systématiquement des différences interaurales de temps, de niveau et spectrales.

Dans ce contexte, le projet ANR LoudNat (Loudness of Natural Sounds) est un consortium entre le Laboratoire Vibrations Acoustique (LVA, INSA de Lyon), l'équipe Sons du Laboratoire de Mécanique et d'Acoustique (LMA, CNRS de Marseille) et l'équipe Perception et Design Sonore (PDS) de l'Institut de Recherche et de Coordination Acoustique/Musique (IRCAM, Paris). Il s'articule en trois phases d'étude autour de la sonie, dans le but de mieux comprendre et prédire la sonie pour

des sons "naturels". Tandis qu'une première partie du projet concerne l'influence des propriétés instationnaires des sons environnementaux sur la sonie (thèse de E. Ponsot), une seconde partie s'est focalisée sur l'aspect binaural de la perception de la sonie, pour des sons stationnaires (objet de la thèse présentée). Finalement, la troisième phase vise à combiner les résultats des deux premières et à étudier la sonie d'un champ acoustique non-stationnaire en présence de différences interaurales (source dont la position et/ou le niveau varie au cours du temps).

Le travail de thèse s'intéresse donc à la sonie des sons stationnaires dans des situations d'écoute dichotiques réalistes et a pour but d'apporter de nouveaux éléments dans la compréhension de la formation du percept de sonie globale.

Il est admis qu'un son présenté aux deux oreilles est perçu plus fort que lorsqu'il est présenté à une seule oreille. Ce processus d'intégration est défini comme la sommation de sonie binaurale, quantifiée par le rapport de sonie binaural-à-monaurale (rapport entre la sonie d'un son en présentation diotique et la sonie de ce même son en présentation monaurale). Certaines études ont rapporté un rapport de sonie binaural-à-monaural égal à 2, correspondant à une sommation parfaite de la sonie à chaque oreille (Hellman et Zwislocki (1963), Marks (1978)). Cependant, d'autres études plus récentes contredisent ces résultats, rapportant des valeurs inférieures à 2 (e.g. 1.3 : Marozeau *et al.* (2006), 1.5 : Zwicker et Zwicker (1991), 1.7 : Scharf et Fishken (1970)), traduisant une sommation de sonie "moins que parfaite". Notamment, une série d'études récentes sur la sonie d'une source artificielle spatialisée en champ libre a montré qu'un gain binaural de 3 dB (correspondant approximativement à un rapport binaural-à-monaural de 1.2) était optimal pour prédire la sonie d'une source dans un environnement réel, malgré l'importance des différences interindividuelles (Sivonen *et al.* (2005), Sivonen et Ellermeier (2006), Sivonen (2007), Sivonen et Ellermeier (2008)).

La première partie de la thèse s'inscrit dans la lignée de cette série d'études, en essayant de comprendre dans quelle mesure la variabilité interindividuelle observée par les auteurs peut être expliquée par une variabilité dans les pentes des fonctions de sonie monaurales. Dans une deuxième partie, la sommation de sonie binaurale est étudiée pour des champs acoustiques plus complexes, en présence d'une ou plusieurs sources artificielles ou réelles.

D'autre part, quelques modèles de sonie binaurale existent. Le modèle ANSI-S3.4 (2007) permet d'estimer la sonie à partir d'un enregistrement binaural sur une tête artificielle, mais reste basé sur une sommation de sonie parfaite. En revanche, deux modèles de sonie binaurale ont été récemment proposés, permettant de rendre compte d'une sommation de sonie moins que parfaite. Le premier est le modèle de Moore et Glasberg (2007), proposant des modifications de l'ANSI-S3.4 (2007) pour prendre en compte les inhibitions contralatérales (Gigerenzer et Strube, 1983). Le deuxième est le modèle de Sivonen et Ellermeier (2008) basé sur une règle de sommation des niveaux de pression aux tympans introduite par Robinson et Whittle (1960) et ajustée à partir des nombreuses données subjectives recueillies par les auteurs.

Ainsi, tout au long de cette étude, les données subjectives recueillies auprès d'auditeurs ont été confrontées à ces différents modèles afin d'évaluer leur performance dans des situations d'écoute de plus en plus réalistes.

L'ensemble du travail est finalement construit autour des questions générales suivantes : Dans l'hypothèse où la sonie peut être prédite à partir des signaux reçus aux oreilles de l'auditeur, comment ces signaux (à chaque oreille) sont-ils combinés pour aboutir à un unique percept de sonie binaurale ?

Dans un environnement réel, dans quelle mesure la sommation de sonie binaurale peut-elle varier d'un auditeur à l'autre et quelles en sont les raisons ? Les modèles de sonie existants permettent-ils de prédire la sonie de champs acoustiques stationnaires complexes en présence de sources réelles ? En pratique, comment prédire la sonie globale à partir de mesures binaurales réalisées sur un mannequin acoustique dans des situations d'écoute fortement dichotiques ?

Organisation de l'étude :

Chapitre 1 Le premier chapitre présente le contexte de l'étude concernant principalement : la sommation de sonie binaurale, la sonie binaurale des champs acoustiques et les modèles de sonie binaurale. Le document s'articule ensuite en deux parties principales.

Première partie Sonie directionnelle et variabilité interindividuelle

Chapitre 2 Le deuxième chapitre a consisté à reproduire l'expérience de Sivonen et Ellermeier (2008) sur un plus grand nombre d'auditeurs et pour un plus grand nombre d'angles d'incidence. Le but principal était de mieux quantifier et comprendre l'origine des différences interindividuelles importantes qui avaient été rapportées par les auteurs. L'expérience consistait à mesurer l'influence de l'angle d'incidence sur la sonie pour des bandes de bruit tiers d'octave (centrées à 1 kHz et 5 kHz), en comparant la sonie d'une source spatialisée dans différentes directions avec la sonie d'une source de référence frontale. Les sources étaient simulées au casque, via un processus de synthèse binaurale générique. Le modèle proposé par Sivonen et Ellermeier (2008) a ensuite été ajusté à partir des données expérimentales afin d'estimer les gains de sommation binaurale individuels et médians.

Chapitre 3 Le troisième chapitre propose quant à lui d'explorer une hypothèse pour expliquer la variabilité dans la sommation de sonie binaurale entre les auditeurs. Cette hypothèse est l'implication de la variabilité des fonctions de sonie monaurales individuelles dans la variabilité des sensibilités directionnelles en sonie. Ainsi, les fonctions de sonie monaurales pour des bandes de bruit tiers d'octave centrées à 1 kHz et à 5 kHz ont été mesurées pour quatre auditeurs à l'aide de deux méthodes : une estimation de grandeur (ME) et une production de rapport (RP). La variabilité théorique induite par les différences d'exposants individuels dans la sensibilité directionnelle en sonie a finalement été comparée à la variabilité totale mesurée dans le premier chapitre.

Deuxième partie Sonie binaurale dans un champ acoustique stationnaire complexe

Chapitre 4 Le quatrième chapitre est centré sur l'étude d'un champ acoustique multi-tonal, composé d'une source bitonale (deux bandes de tiers d'octave) ou de deux sources tonales spatialement distribuées. Il s'inscrit dans le prolongement d'une étude menée par Song (2007). A partir d'une même méthodologie générale, les objectifs (et donc les conditions expérimentales) ont été déclinés en trois grandes parties. Une première partie est consacrée à l'étude de la sonie directionnelle d'une source de bruit à deux composantes tonales (centrées à 1 kHz et 5 kHz), dans le plan horizontal. La deuxième partie porte sur l'influence de la distribution spatiale de deux sources tonales simultanées sur la sonie, pour des grands angles d'incidence (60° et 90°). Dans la troisième partie, la sonie de la

source bitonale d'incidence frontale est finalement comparée à la sonie d'une simple source tonale. La performance des modèles de sonie binaurale existants est discutée pour l'ensemble des conditions expérimentales.

Chapitre 5 Afin de tester la validité des modèles de sonie dans des situations plus "naturelles" d'écoute, le cinquième chapitre propose finalement de mesurer la sonie pour des sources réelles dans une acoustique réelle. Ainsi des mesures ont été réalisées pour deux types de sources spatialisées (aspirateur, machine à laver) ou pour un champ acoustique constitué d'une multitude de sources (ventilation de voiture, ou voiture en soufflerie).

L'expérience présentée dans le chapitre 5 a été réalisée à l'Institut für psychologie, à la Technische Universität de Darmstadt, dans le cadre d'un échange de 3 mois dans le laboratoire du professeur W. Ellermeier (financé par le LabEx CeLyA).

Contexte de l'étude

1.1 Sonie des sons stationnaires : principes fondamentaux

1.1.1 Définitions

Par définition, la *sonie* est une grandeur perceptive qui désigne la force sonore perçue d'un son. La sonie est exprimée en sones, échelle numérique proportionnelle à la sensation perçue. Un sone correspond à la sensation produite par un sinus à 1000 Hz à un niveau de pression de 40 dB SPL. Un doublement de la valeur en sone correspond à un doublement de la sensation. Ainsi, un son de 2 sones sera perçu deux fois plus fort qu'un son de 1 sone et ainsi de suite. Pour des sons non stationnaires ou en présence de sons complexes, la sonie peut varier au cours du temps ou selon la partie du son sur laquelle l'auditeur porte son jugement. On définira alors la *sonie globale* comme la sensation du niveau perçu d'un son dans son ensemble.

La sonie dépend principalement du niveau de pression acoustique du stimulus, mais dépend également de ses caractéristiques spectrales, de sa durée, ou de son enveloppe temporelle (dans le cas de sons non-stationnaires). Ainsi, deux sons qui ont la même intensité, mais qui diffèrent suivant d'autres paramètres physiques, aboutiront vraisemblablement à des sensations de sonie différentes. L'influence des principaux facteurs sur la sonie est abordée dans la partie 1.1.2.

Dans de nombreux cas, l'évaluation de la sonie d'un son implique la comparaison avec la sonie d'un autre son ayant des caractéristiques physiques différentes. La sensation de sonie peut alors être exprimée en termes de niveau SPL du son de comparaison à sonie égale. Si le son de comparaison est un sinus à 1 kHz (en champ libre, en incidence frontale), son niveau SPL correspond au *niveau d'isotonie*. L'unité du niveau d'isotonie est le phone. Un niveau de N phones correspond à la sensation de sonie produite par un sinus de fréquence 1 kHz à un niveau de pression de N dB SPL. Le niveau d'isotonie permet finalement d'exprimer le rapport de sonie entre deux sons en terme de gain équivalent qu'il aurait fallu appliquer à un sinus à 1 kHz pour avoir la même différence de sensation (référence parfois plus représentative qu'un rapport de sonie).

1.1.2 Les principaux facteurs qui influent sur la sonie

La sonie varie donc en fonction de l'intensité du stimulus, mais également en fonction de l'ensemble de ses différents attributs physiques (intensité, fréquence, largeur de bande, durée, présence d'autres sons,...). Autrement dit, il n'existe pas de correspondance exclusive entre la sonie et un paramètre physique du stimulus en particulier, quel qu'il soit. On notera que la sonie peut également varier en fonction d'autres facteurs qui ne dépendent pas du stimulus comme le contexte, l'auditeur, etc... (cf. § 1.3.4). On se contentera cependant ici de rappeler comment la sonie varie selon les principales dimensions physiques du stimulus : l'intensité, la fréquence, la largeur de bande et la durée.

1.1.2.1 L'intensité

La sonie croît de manière monotone avec l'intensité du stimulus, dans les limites de la plage de dynamique du système auditif d'environ 120 dB. La fonction psychophysique qui relie l'intensité (ou la pression) du stimulus à la sonie est appelée fonction de sonie. Le modèle de fonction de sonie le plus simple et couramment utilisé a été établi par Stevens (1955) pour un son pur à 1 kHz, à l'aide de plusieurs méthodes d'estimation directe. Ce modèle définit la relation entre la sonie (S) en sones et la pression (P) en pascals pour un son pur à 1 kHz, par la loi de puissance décrite dans l'équation 1.1 (voir aussi, figure 1.1).

$$S = k.P^\alpha \quad (1.1)$$

avec $\alpha = 0.6$ et $k = 0.01$ (si P est exprimé en μPa).

Les différentes études de la littérature s'accordent sur le fait qu'une fonction de puissance fournit une bonne approximation de la fonction psychométrique pour les niveaux supérieurs à 40 dB SPL. Plusieurs coefficients ont été proposés, mais la valeur $\alpha = 0.6$ est souvent utilisée. Pour cette valeur, une augmentation du niveau de pression de 10 dB correspond approximativement à un doublement de la sonie.

Pour les niveaux faibles (entre le seuil d'audition et 40 dB SPL environ), la pente de la fonction de sonie est plus importante que pour les niveaux moyens (voir Buus *et al.* (1998) pour une revue). De nombreuses propositions de modélisation de la fonction de sonie pour les niveaux inférieurs à 40 dB SPL ont été avancées. Les différents modèles fournissent des prédictions relativement proches. Une des principales différences réside dans la prédiction de la sonie au seuil, considérée égale à zéro (e.g. Scharf (1978), Zwislocki et Hellman (1960)) ou non (e.g. Zwislocki (1965), Buus *et al.* (1998)), suivant les études. Cette question de la valeur de la sonie au seuil d'audition est toujours d'actualité. A titre d'exemple, l'équation 1.2 décrit la fonction de sonie adoptée dans la norme ISO-226 (2003) pour interpoler les niveaux d'isosonie, basée sur une suggestion de Zwislocki et Hellman (1960) et Lochner et Burger (1961).

$$S = c.(p^{0.5} - p_{th}^{0.5}) \quad (1.2)$$

où p est la pression du stimulus dont on veut estimer la sonie et p_{th} est la pression au seuil d'audition (fixé à 2.4 dB SPL à 1 kHz). Selon ce modèle, la sonie est par exemple nulle au seuil d'audition. Actuellement, différentes fonctions de sonie tenant compte de l'augmentation de la pente aux bas niveaux sont utilisées dans les normes ISO-226 (2003) (eq. 1.2), ISO-532B (1975) et ANSI-S3.4 (2007). Les fonctions issues de ces trois modèles pour un son pur à 1 kHz sont comparées dans la figure 1.1.

De la même manière, plusieurs études s'accordent à décrire une augmentation de la pente des fonctions de sonie aux forts niveaux (e.g. Viemeister et Bacon (1988), Buus *et al.* (1997), Florentine *et al.* (1996)). Cependant, les niveaux supérieurs à 100 dB SPL ont été très peu étudiés, étant donnés les risques pour la santé qu'implique la restitution de sons purs à de tels niveaux.

L'augmentation plus rapide de la sonie avec le niveau aux faibles et forts niveaux, a été mise en avant dans des études récentes sur l'intégration temporelle (e.g. Florentine *et al.* (1996), Buus *et al.* (1997), Epstein et Florentine (2005)). Ces articles supportent l'hypothèse suivant laquelle pour les sons courts (jusqu'à quelques centaines de millisecondes), une même augmentation de la durée induit

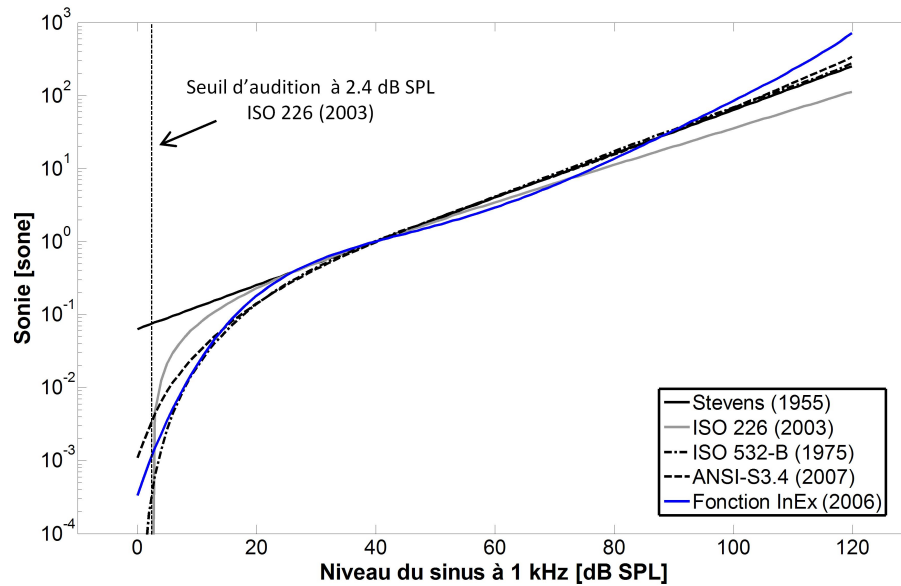


FIGURE 1.1: Fonctions de sonie pour un son pur de fréquence 1 kHz. Comparaison des fonctions telles qu'elles sont définies dans certains standards de sonie actuels, avec la loi de puissance de Stevens (1955) ou encore la fonction InEx proposée plus récemment par Florentine et Epstein (2006).

une augmentation de la sonie par un même facteur, quel que soit le niveau des stimuli (Florentine *et al.*, 1996). Cependant, lorsque le niveau d'un son pur de 5 ms est ajusté au point d'égalité subjectif (PSE) en sonie avec un son pur de 200 ms, le niveau du son de 5 ms est plus élevé, mais la différence de niveau est plus grande aux intensités moyennes. Ainsi, d'après l'hypothèse d'un rapport de sonie constant avec le niveau, la plus grande quantité d'intégration temporelle aux niveaux moyens reflète une pente moins importante de la fonction de sonie. Afin de rendre compte de cet effet de compression de la sonie aux niveaux moyens, des modifications de la loi de puissance standard ont alors été proposées par Buus et Florentine (2002) puis définies par Florentine et Epstein (2006) sous l'appellation "inflected exponential law" (ou fonction InEx). Cette fonction de sonie définie par un polynôme du 5^e degré, admet une pente proche de 2 aux faibles niveaux, puis proche de 0.6 entre 40 et 60 dB SPL, et augmente à nouveau progressivement aux forts niveaux. La fonction InEx est également représentée dans la figure 1.1. Une fonction de sonie similaire a été mesurée par Marozeau *et al.* (2006) ou dérivée des résultats de Whilby *et al.* (2006) dans des études analogues pour évaluer la sommation de sonie binaurale (cf. § 1.2.1.1).

La plupart du temps, la fonction de sonie au-dessus de 40 dB SPL reste néanmoins approchée par une loi de puissance (Stevens, 1955) dont on fait varier l'exposant. De nombreuses études ont montré que l'exposant de la fonction de sonie pouvait varier de manière importante entre deux auditeurs (e.g., Stevens (1955), Hellman et Meiselman (1988), Hellman et Meiselman (1993)), mais la plupart semblent s'accorder sur une valeur moyenne de l'exposant autour de 0.6 à 1 kHz. Pour un son pur à 3 kHz, la valeur $\alpha = 0.67$ est habituellement admise (Stevens, 1972).

1.1.2.2 La fréquence

Pour un son pur, la sonie varie également en fonction de sa fréquence. Les courbes de niveau d'isotonie décrivent la variation de niveau nécessaire pour conserver la même sonie en fonction de la fréquence, sur l'ensemble de la plage de niveaux. Par convention, le niveau d'isotonie est exprimé en phones, correspondant au niveau SPL d'un son pur à 1 kHz. La figure 1.2 représente les courbes d'isotonie décrites dans la norme ISO-226 (2003). Tous les points d'une même courbe produisent la même sonie (un son à 50 dB SPL à 250 Hz aura par exemple la même sonie qu'un son à 40 dB SPL à 1 kHz). La courbe en pointillé représente le seuil d'audition moyen (en champ libre, incidence frontale).

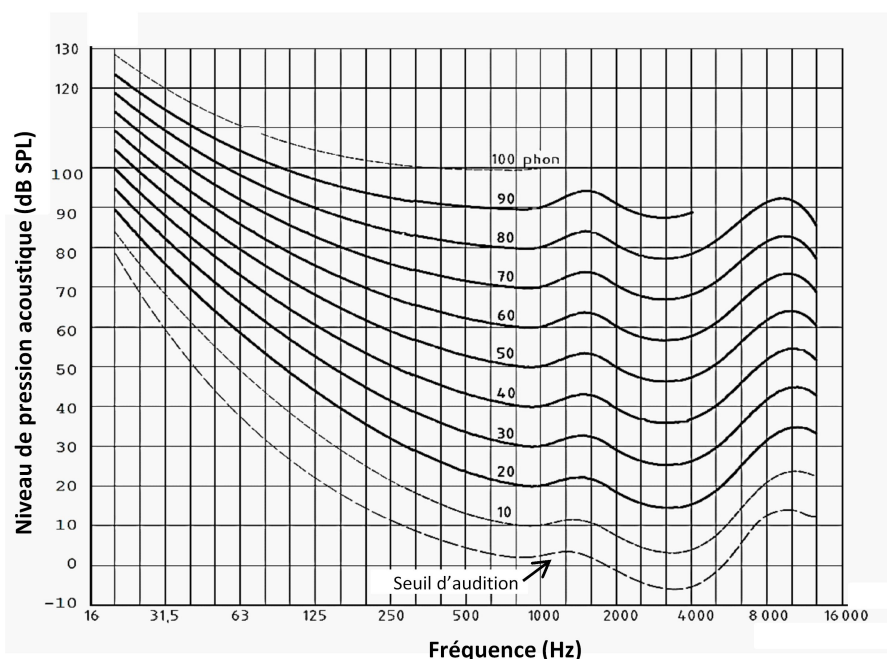


FIGURE 1.2: Courbes de niveau d'isotonie issue du standard ISO-226 (2003).

Les niveaux d'isotonie sont mesurés à l'aide de méthodes d'ajustement ou de comparaison par paires, les stimuli sont restitués en incidence frontale en champ libre (à l'exception de quelques études, dont notamment les premières courbes de référence de Fletcher et Munson (1933), qui avaient été mesurées au casque). Les courbes de Fletcher et Munson (1933) ont été très utilisées depuis des années, au même titre que les courbes de Robinson et Dadson (1956) ayant fait l'objet des normes ISO/R-226 (1961) puis ISO-226 (1987). Mais les études plus récentes (e.g. Suzuki *et al.* (1989), Watanabe et Møller (1990), Takeshima *et al.* (2001)) qui ont mené à la révision de la norme ISO de 2003, ont rapporté des différences parfois importantes concernant la contribution des basses fréquences (en-dessous de 800 Hz), qui était vraisemblablement surestimée dans les courbes de Robinson et Dadson (1956). A titre d'exemple, la figure 1.3 tirée de l'étude de Suzuki et Takeshima (2004) propose une comparaison des résultats obtenus par les différentes études réalisées depuis 1983 pour la courbe à 40 phones, par rapport à la courbe initiale proposée par Robinson et Dadson (1956).

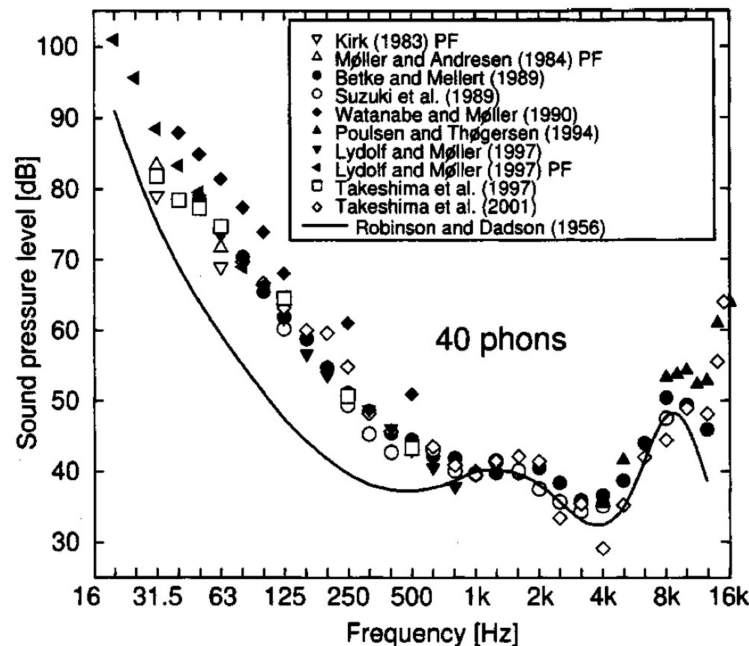


FIGURE 1.3: Fig. 1 de Suzuki et Takeshima (2004). Courbe de niveau d'isophonie à 40 phons pour des sons purs. La ligne continue représente la courbe mesurée par Robinson et Dadson (1956), adaptée comme standard international ISO/R-226 (1961) et ISO-226 (1987). Les symboles représentent les données expérimentales collectées depuis 1983.

On notera que les niveaux d'isophonie dérivés des modèles de sonie standardisés DIN-45631 (1990) (issu du modèle ISO-532B (1975)) et ANSI-S3.4 (2007) diffèrent un peu de ceux décrits dans la norme ISO-226 (2003), le modèle ANSI-S3.4 (2007) étant globalement le plus proche (Charbonneau *et al.*, 2009). Il est possible que cette différence soit due à un ajustement des modèles basé sur les anciens standards d'isophonie (ISO-226, 1987). La norme ISO-226 (2003) fournit probablement une prédiction plus précise de la sonie en fonction de la fréquence (à niveaux fixes), ayant été ajustée à partir d'un grand nombre de données expérimentales comparant la sonie de sons purs à des fréquences différentes. En revanche, les modèles prédisent vraisemblablement mieux la variation de la sonie en fonction du niveau pour une fréquence donnée.

1.1.2.3 La largeur de bande (intégration spectrale)

Pour un son pur, la sonie varie en fonction de la fréquence. Pour un son complexe, la sonie varie en fonction de la distribution de l'énergie dans la gamme de fréquences audibles. En effet, la manière dont cette énergie est distribuée détermine la manière dont le système auditif va intégrer les différentes composantes fréquentielles du stimulus, et donc le percept de sonie final. Ce processus d'*intégration spectrale* (ou *sommation de sonie*), est directement lié à la notion de bandes critiques, utilisées pour modéliser le mécanisme de filtrage fréquentiel de l'oreille interne.

Si l'ensemble des composantes fréquentielles d'un son complexe est contenu à l'intérieur d'une bande de fréquence inférieure à une bande critique, le logarithme de la sonie globale est proportionnel

au niveau total d'énergie contenue dans la bande critique (somme en énergie). Mais si deux composantes sont séparées par un intervalle en fréquence supérieur à une bande critique, la sonie globale est égale à la somme des sonies de chaque composante (somme en sonie). Ainsi, pour un son complexe à énergie constante, tant que la largeur de bande totale est inférieure à une bande critique, la sonie reste constante. En revanche, au-delà de la bande critique, la sonie augmente avec la largeur de bande (e.g. Zwicker *et al.* (1957), Florentine *et al.* (1978), Leibold *et al.* (2007)). Finalement, il est admis par la plupart des modèles que suite au filtrage fréquentiel de l'oreille interne, une sonie spécifique est calculée dans chaque bande critique puis sommée pour obtenir la sonie globale (Zwicker et Fastl (1999), Moore *et al.* (1997)).

On notera également que la sommation de sonie dépend du niveau, étant plus importante aux niveaux moyens (et quasiment nulle à 20 phons) et que la largeur des bandes critiques augmente avec la fréquence centrale (e.g. Zwicker *et al.* (1957), Zwicker et Fastl (1999)). Plusieurs études décrivent la forme des filtres et la manière dont elle varie avec le niveau (e.g. Greenwood (1961), Glasberg et Moore (1990)).

Actuellement, deux principales représentations de l'organisation des filtres auditifs sont couramment utilisées. La première est l'échelle des Barks définissant 24 bandes critiques adjacentes correspondant à une distance constante d'environ 1.3 mm sur la membrane basilaire et décrivant la plupart du spectre audible (Zwicker, 1961). La largeur des bandes critiques est d'environ 100 Hz en-dessous de 500 Hz puis augmente avec la fréquence. L'ERB (Equivalent Rectangular Bandwidth) est une autre représentation introduite par Moore et Glasberg (1983) dans leur révision du modèle de Zwicker, qui pourra être préférée pour certaines applications. Une ERB correspond à la largeur de bande d'un filtre rectangulaire ayant le même pic de transmission et laissant passer la même énergie pour un bruit blanc en entrée que le filtre auditif à la fréquence considérée. Cette représentation a été introduite pour permettre notamment de mieux rendre compte de l'allure des niveaux d'isosonie en basse fréquence et de la sonie des sons partiellement masqués. Ces notions seront à nouveau introduites dans la partie 1.4.1.

1.1.2.4 La durée (intégration temporelle)

Les courbes d'isosonie et le seuil d'audition présentés dans la figure 1.2 ont été établis pour des stimuli dont la durée était de l'ordre de la seconde. Dans ce cas, la durée n'a pas d'influence sur la sonie.

Cependant, pour des stimuli de durée courte, la sonie augmente avec la durée (à énergie constante) jusqu'à une certaine durée critique, qui varie selon les études entre 15 ms (Small Jr *et al.*, 1962) et 500 ms (Ekman *et al.* (1966), Boullet *et al.* (2004)). C'est ce qu'on appelle l'*intégration temporelle*. La valeur critique varie en fonction de la fréquence et de l'intensité : elle a tendance à diminuer lorsque la fréquence ou le niveau augmente (e.g. Stephens (1974)). Néanmoins, cette dépendance n'est pas mesurée systématiquement pour des niveaux supraliminaires.

Une série d'études récentes s'est attachée à mesurer l'effet de la durée sur la sonie pour des sons purs, à différents niveaux et en diversifiant les méthodes de mesures (Buus *et al.* (1997), Epstein et Florentine (2005)). Ces travaux ont permis d'appuyer l'hypothèse initialement formulée par Florentine *et al.* (1996) selon laquelle, quel que soit le niveau global du stimulus, une même variation de durée induit une même variation de sonie (cf. § 1.1.2.1).

D'autre part, le seuil d'audition est également plus élevé pour des durées inférieures à la durée

critique et décroîtrait d'environ 3 dB par doublement de la durée selon Zwislocki (1960).

Au-delà de la durée critique, à énergie fixe la sonie reste constante, bien que dans certains cas, la sonie puisse commencer à décroître au cours du temps (seulement au bout de plusieurs secondes). On parle d'adaptation simple en sonie, cet effet intervenant principalement pour les fréquences élevées (Miskiewicz *et al.*, 1993) et à des niveaux faibles, proches du seuil d'audition (Hellman *et al.*, 1997). Pour des fréquences supérieures à 10 kHz, l'effet d'adaptation peut néanmoins être marqué même à des niveaux élevés (Miskiewicz *et al.*, 1993).

Une revue plus complète de l'effet de la durée sur la sonie a été réalisée par Boullet (2005).

1.2 Sonie binaurale

Une seule oreille est nécessaire pour qu'un percept de sonie soit créé, et les différents facteurs influant sur la sonie présentés dans la partie 1.1.2 s'appliquent à la sonie d'une seule oreille. Cependant, dans un environnement réel, nous utilisons constamment nos deux oreilles (il est d'ailleurs impossible de "fermer une oreille" comme nous fermons un œil) et il est donc important de comprendre comment les percepts de sonie correspondant à chaque oreille sont combinés en un unique percept de sonie binaurale.

1.2.1 Sommation de sonie binaurale

Il est admis qu'un même son présenté aux deux oreilles est perçu plus fort que s'il est présenté à une seule oreille (e.g. Fletcher et Munson (1933), Marks (1978)). Cette interaction constructive d'un point de vue perceptif est désignée par le terme de *sommation de sonie binaurale*.

La plupart des études sur la sommation de sonie binaurale ont été réalisées au casque, afin de pouvoir mieux contrôler les stimuli présentés à chaque oreille, indépendamment des différentes transformations acoustiques qui opèrent dans un environnement réel. Différentes configurations de présentation des sons peuvent alors être envisagées : une présentation monaurale (présentation des sons à une seule oreille), ou une présentation binaurale pouvant être soit diotique (un même signal est présenté aux deux oreilles) soit dichotique (des signaux différents sont présentés à chaque oreille).

On notera que cette partie sur la sommation de sonie binaurale est largement inspirée de la revue complète sur ce sujet déjà réalisée par Sivonen et Ellermeier (2011).

1.2.1.1 Comment peut-on quantifier la sommation de sonie binaurale ?

La sommation de sonie binaurale est évaluée principalement à travers deux quantités : 1) le *gain de sommation binaurale* ou *gain binaural*, correspondant au gain à appliquer à un signal présenté en monaural pour qu'il produise la même sonie que ce même signal présenté de manière diotique 2) le *rapport de sonie binaural-à-maural*, correspondant au rapport entre la sonie d'un son présenté de manière diotique et la sonie de ce même son présenté en monaural. Ces deux quantités sont directement reliées par la pente de la fonction de sonie, et peuvent être déduites directement l'une de l'autre uniquement si les pentes des fonctions de sonie monaurale et binaurale sont constantes et égales. En effet, si on considère la fonction de sonie comme une simple fonction de puissance

(Stevens, 1955), les fonctions de sonie monaurale et binaurale sont décrites respectivement par les équations 1.3 et 1.4.

$$S_m = k_m \cdot p_m^{\alpha_m} \quad (1.3)$$

$$S_b = k_b \cdot p_b^{\alpha_b} \quad (1.4)$$

En faisant l'hypothèse que le rapport de sonie binaural à monaural S_b/S_m est constant avec le niveau (BELRH, pour Binaural equal-loudness ratio hypothesis, voir par exemple Fletcher et Munson (1933), Marozeau *et al.* (2006)), les fonctions de sonie monaurale et binaurale ont alors la même pente ($\alpha_m = \alpha_b = \alpha$), et le gain de sommation binaurale g peut être exprimé simplement en fonction du rapport de sonie binaural à monaural S_b/S_m (eq. 1.5).

$$g = \frac{20 * \log_{10}(S_b/S_m)}{\alpha} \quad (1.5)$$

En d'autres termes, si la sonie est représentée sur une échelle logarithmique en fonction du niveau SPL, le décalage horizontal entre les deux courbes représente le gain binaural, tandis que le décalage vertical représente le rapport de sonie binaural-à-monaural (cf. figure 1.4). Un rapport de sonie binaural-à-monaural égal à 2 traduit une sommation de sonie dite "parfaite" (un son présenté en binaural est perçu deux fois plus fort que le même son présenté en monaural, les sonies correspondant à chaque oreille se somment). En accord avec l'équation 1.5, le gain de sommation binaurale correspondant dépend de l'exposant de la fonction de sonie.

Selon les paramètres qui sont fixés et les paramètre qui sont mesurés, il est possible d'arriver à des résultats qui semblent contradictoire. Par exemple, si dans une première étude les fonctions de sonie ont un coefficient directeur $\alpha = 0.6$ (figure 1.4, courbes noires), si un gain de sommation binaurale de 10 dB est mesuré, cela correspond à un rapport S_b/S_m égal à 2. Cependant, dans des conditions expérimentales différentes où la fonction de sonie a un coefficient directeur $\alpha = 0.4$ (figure 1.4, courbes grises), si un gain de sommation binaurale de 15 dB est mesuré, malgré l'apparente contradiction avec l'étude précédente, cela correspond à nouveau à un rapport binaural-à-monaural égal à 2. Il est ainsi possible de rapprocher des résultats a priori contradictoires entre certaines études de la littérature en se ramenant à un coefficient commun pour la fonction de sonie ou à un rapport de sonie binaural-à-monaural commun (exemples dans Marks (1978) ou Whilby *et al.* (2006)).

Le gain de sommation binaurale est mesuré principalement par des méthodes de comparaison de sonie (ajustement, procédures adaptatives,...). La plupart des études rapportent un gain binaural plus faible près du seuil, qui augmente jusqu'à environ 40 dB SL (e.g. Fletcher et Munson (1933), Caussé et Chavasse (1942), Marks (1978)). Les gains binauraux maximaux obtenus pour des sons purs sont globalement compris entre 5 dB (Scharf et Fishken, 1970) et 10 dB (Marks, 1978). Des gains du même ordre de grandeur ont été obtenus pour des bruits de différentes largeurs de bande (e.g. Irwin (1965), Reynolds et Stevens (1960)).

Contrairement au gain de sommation binaurale, le rapport de sonie binaural-à-monaural est estimé dans la majorité des cas à partir de la différence entre les fonctions de sonie monaurales et binaurales, elles-mêmes estimées à l'aide de méthodes de jugement direct de la sonie (estimation de

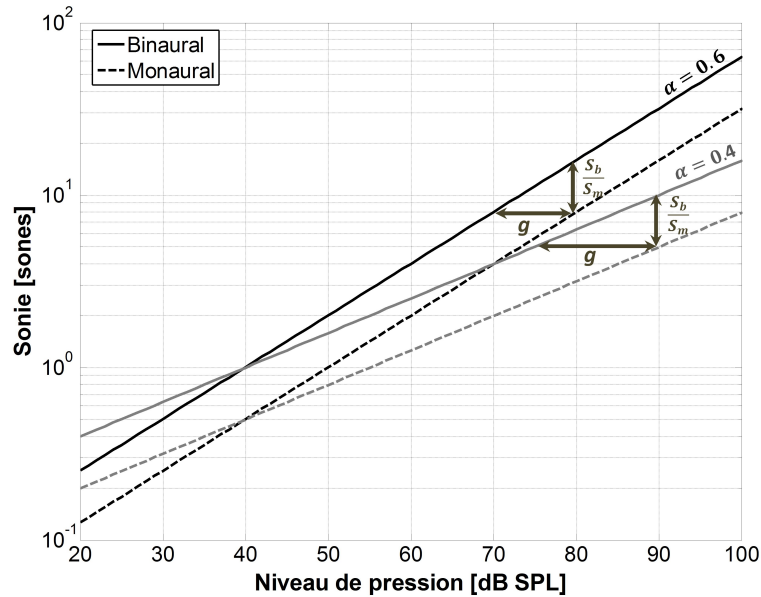


FIGURE 1.4: Fonctions de sonie monaurales (courbes en pointillés) et binaurales (courbes continues) de la forme $S = k \cdot P^\alpha$ pour $\alpha = 0.4$ (en gris) et $\alpha = 0.6$ (en noir). La valeur de k est ajustée pour que les fonctions de sonie binaurales passent par le point (40 dB SPL ; 1 sone). On constate qu'à un même rapport de sonie binaural-à-monastral ($S_b/S_m = 2$), peut correspondre plusieurs valeurs du gain binaural g en fonction de la valeur de la pente de la fonction psychométrique α .

grandeur, production de grandeur, appariement inter-modal (CMM),...). Un certain nombre d'études s'accordent à dire que le rapport de sonie binaural-à-monastral est égal à 2, correspondant ainsi à une sommation de sonie parfaite (e.g. Hellman et Zwislocki (1963), Marks (1978)). Cependant, d'autres études plus récentes contredisent cette théorie, rapportant des valeurs inférieures à 2 (e.g. 1.3 : Marozeau *et al.* (2006), 1.5 : Algorn *et al.* (1989b) Marks (1980), Zwicker et Zwicker (1991), 1.7 : Scharf et Fishken (1970), etc.), ce qui traduit donc une sommation de sonie "moins que parfaite".

Une hypothèse avancée pour expliquer ces différences est que le recours à des méthodes de jugement direct de la sonie peut impliquer entre autre, une déviation entre l'échelle de nombres utilisée par l'auditeur et l'échelle de sensation (de sonie) "réelle" (Ellermeier et Faulhammer (2000), DeCarlo (2005)). La mesure du gain binaural par une méthode d'égalisation en sonie (entre une situation monaurale et une situation binaurale) semble finalement être l'approche la moins biaisée (impliquant le moins d'hypothèses). Néanmoins, comme nous l'avons vu précédemment, le gain binaural ne permet pas d'obtenir l'information en terme de sonie si l'allure de la fonction psychophysique n'est pas connue (éq. 1.5). Par exemple, si le gain binaural varie en fonction du niveau, il est difficile de savoir si cette variation correspond à une variation de la pente de la fonction de sonie, à une variation du rapport de sonie binaural-à-monastral ou à une variation combinée des deux paramètres.

Deux études récentes complémentaires ont apporté des éléments de compréhension en combinant les deux méthodes. Dans Whilby *et al.* (2006) le gain de sommation binaurale a été mesuré

entre 10 et 90 dB SPL pour des sons purs de 5 et 200 ms à 1 kHz. Les gains binauraux obtenus étaient similaires pour les deux durées, mais variaient en fonction du niveau, étant plus élevés aux niveaux moyens (entre 50 et 70 dB SPL). En partant de l'hypothèse d'un rapport de sonie binaural-à-monastral constant avec le niveau (BELRH), les auteurs ont proposé un modèle permettant de dériver les fonctions de sonie à partir du gain binaural, pour un rapport de sonie binaural-à-monastral fixé. Comme le gain binaural variait en fonction du niveau, cela impliquait donc que la pente de la fonction de sonie variait également en fonction du niveau (pente plus faible aux niveaux moyens). Une étude complémentaire de Marozeau *et al.* (2006) a consisté à mesurer les fonctions de sonie monaurale et binaurale pour les mêmes stimuli (à 200 ms), à l'aide d'une méthode d'appariement inter-modal (CMM) de la longueur d'une corde à la sonie. Tous les niveaux compris entre le seuil d'audition et 100 dB SPL ont été testés par pas de 5 dB. Les fonctions de sonie obtenues avaient à nouveau une forme en "S" (pente plus faible aux niveaux moyens), résultat conforté par la coïncidence de la fonction de sonie monaurale avec celle qui avait été rapportée par Epstein et Florentine (2005) pour le même stimulus dans le cadre d'une série d'études sur l'intégration temporelle. D'autre part, le rapport de sonie binaural-à-monastral (estimé à partir des fonctions de sonie monaurale et binaurale) était égal à 1.3 et constant avec le niveau, confirmant ainsi l'hypothèse de départ (BELRH) qui avait été admise dans Whilby *et al.* (2006). La figure 1.5 tirée de Marozeau *et al.* (2006) résume ces résultats.

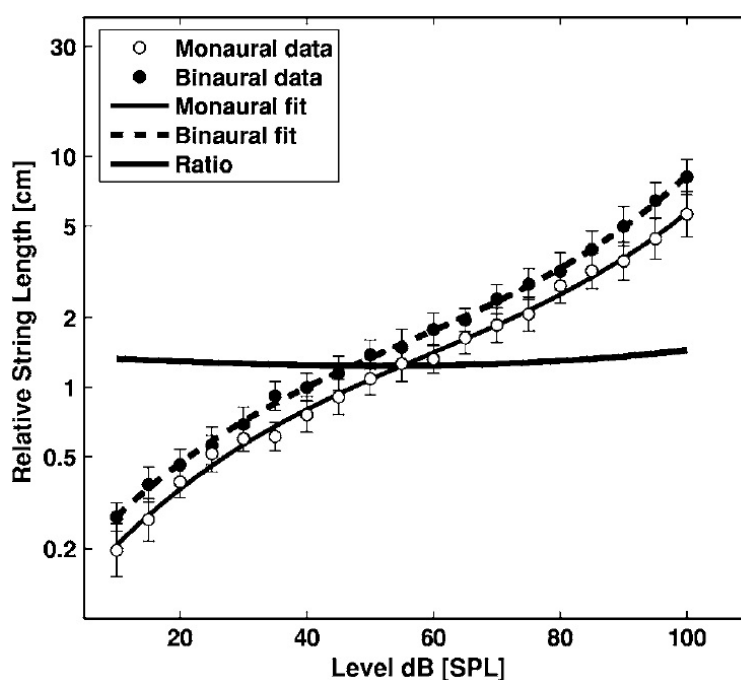


FIGURE 1.5: Fig. 2 de Marozeau *et al.* (2006). Données moyennes obtenues pour 8 auditeurs. Les polynômes d'ordre trois ajustés aux données ont été normalisés en divisant chaque longueur de corde par la longueur moyenne obtenue à 40 dB SPL en binaural. Les barres d'erreurs verticales représentent \pm l'erreur standard du logarithme de la longueur de corde.

Finalement, en considérant un rapport binaural-à-monastral fixe de 1.3, les fonctions de sonie

monaurale et binaurale dérivées des égalisations en sonie de Whilby *et al.* (2006) coïncident avec les fonctions de sonie mesurées directement par Marozeau *et al.* (2006). La valeur du rapport de sonie de 1.3 (indiquant une sommation de sonie faible) reste cependant soumise à condition que l'échelle des longueurs de corde (variable de réponse) constitue réellement une échelle de sonie.

Une approche originale a également été proposée récemment par Schlittenlacher *et al.* (2014), où le gain de sommation binaurale a été estimé en égalisant les temps de réaction entre les présentations monaurale et binaurale du stimulus (sinus de 200 ms à 1 kHz). Un gain binaural d'environ 5 dB a été mesuré entre 45 et 80 dB SPL, correspondant à un rapport de sonie binaural-à-monastral égal à 1.4 pour une pente de la fonction de sonie $\alpha = 0.6$. Cela soutient, une fois de plus, l'hypothèse d'une sommation de sonie moins que parfaite et rejoint les résultats de Marozeau *et al.* (2006).

1.2.1.2 Facteurs influençant la sommation de sonie binaurale

L'objectif est de lister une partie des études ayant contribué à comprendre l'influence de certains facteurs sur la quantité de sommation de sonie binaurale. Les études à ce sujet sont nombreuses (notamment dans le cas diotique pour des stimuli simples), mais nous avons plutôt choisi de mettre en avant la diversité des facteurs et d'insister sur l'implication des différences interaurales (dans les situations dichotiques d'écoute).

Un certain nombre d'études ont évalué la sommation de sonie binaurale en se rapportant à des situations diotiques d'écoute. Peu de facteurs semblent avoir une influence dans ce cas. Pour des sons purs, il a par exemple été montré qu'il n'y avait pas d'effet ni du niveau (e.g. Marozeau *et al.* (2006)), ni de la durée (Whilby *et al.* (2006), comparaison des durées 5 ms et 200 ms), ni de la fréquence (Marks (1978), comparaison des fréquences 100, 400 et 1000 Hz) sur la quantité de sommation de sonie binaurale. Le recours à des bruits à bande large ou étroite plutôt qu'à des sons purs peut avoir différentes influences sur le rapport de sonie binaural-à-monastral (e.g. Marks (1980), Algom *et al.* (1989b)). Mais ce n'est pas le cas de la forme globale du bruit (rouge, rose ou bleu) ou de la fréquence centrale pour des bandes de bruit tiers d'octave (Zwicker et Zwicker, 1991). On notera que certaines études ont des conclusions parfois différentes, et que l'effet des facteurs mis en évidence n'est pas toujours très clair.

Lorsque des différences interaurales de niveau (ILD) sont introduites, bien que certains auteurs rapportent l'absence d'effet de l'ILD sur la sonie binaurale (e.g. Keen (1972), Zwicker et Zwicker (1991)), selon des études récentes la sommation binaurale semblerait augmenter avec la différence de niveau entre les deux oreilles, effet moins prononcé en hautes fréquences (Zhang et Mao (2010), Ih et Cha (2003)). D'autre part, pour un rapport signal à bruit faible, il est établi qu'un son pur (en basse fréquence) présenté dans un bruit de manière diotique est perçu plus fort en opposition de phase (entre les deux oreilles) qu'en phase (e.g. Hirsh (1948), Townsend et Goldstein (1972)). Cependant, ce phénomène semble d'avantage associé à un effet de "démasquage binaural" qu'à un effet réel de la phase sur la sonie. Pour des rapports signal à bruit plus grands (au moins 20 dB), aucun effet des différences interaurales de phase n'a été trouvé par Pin et Chocholle (1970, 1971) pour des sons purs de fréquence 45, 60 et 400 Hz. En revanche, Mulligan *et al.* (1985) ont mesuré un gain binaural plus important lorsque la différence interaurale de phase entre deux sinusoïdes ($250 \text{ Hz} \leq f \leq 1400 \text{ Hz}$) était inférieure au déphasage maximal théorique en présence d'une source

réelle (correspondant à un angle de 90° d'azimut). Lorsque cette valeur maximale théorique était dépassée, le gain de sommation binaurale chutait d'environ 3 dB. D'autre part une augmentation de la sonie avec les différences interaurales de temps (correspondant à des angles d'azimut compris entre 0 et 90°) a été mis en évidence pour des sons purs de 200 et 400 Hz à 40 phones (Koehl et Paquier, 2014, 2015). Mais aucun effet n'a été relevé par les auteurs au seuil d'audition ou à 70 phones.

Un autre résultat intéressant a été rapporté récemment par Edmonds et Culling (2009) concernant l'effet de la corrélation interaurale sur la sommation de sonie binaurale. Il a été montré que des bruits dont l'énergie est contenue principalement en basses fréquences sont perçus environ 2 dB plus fort lorsque la corrélation interaurale est nulle ($\rho = 0$) que lorsque les signaux à chaque oreille sont corrélés (qu'ils soient en phase ($\rho = 1$) ou en opposition de phase ($\rho = -1$)). Ce résultat a été constaté pour des bandes de bruit de largeur 1 ERB, centrées à des fréquences de 250, 500, 1000 et 2000 Hz. L'effet était fortement réduit pour un bruit de bande plus large à des fréquences plus élevées (2894-4763 Hz) et inexistant pour un bruit large bande (100-5000 Hz). Le fait que la sommation de sonie binaurale ne soit pas dépendante de la corrélation interaurale pour des bruits large-bande confirme les résultats déjà obtenus par Dubrovskii *et al.* (1972) et Eichenlaub *et al.* (1996).

Concernant les différences de fréquence ou spectrales entre les deux oreilles, il n'y a pas a priori d'effet de la séparation en fréquence entre les deux oreilles sur la sonie binaurale pour des sons purs (Scharf, 1969). En revanche, d'après les travaux de Zwicker et Zwicker (1991), la manière de distribuer l'énergie spectrale entre les deux oreilles semble avoir une influence sur la sommation de sonie. Un son large-bande était décomposé en 24 bandes de fréquences consécutives (correspondant à l'échelle des Barks). Ces bandes étaient ensuite regroupées par bandes de fréquence plus larges de 1, 3, 6 ou 12 bandes critiques consécutives, puis ces bandes plus larges étaient distribuées soit à l'oreille droite, soit à l'oreille gauche (par alternance sur l'ensemble du spectre). La sonie augmentait lorsque le nombre de bandes larges diminuait et donc que leur largeur augmentait (maximum de sonie atteint pour deux bandes de 12 Barks).

L'ensemble de ces études s'est intéressé à l'influence d'un grand nombre de configurations des signaux gauche et droit sur le processus de sommation de sonie binaurale, mais dans le cadre d'une présentation des stimuli au casque et donc dans des conditions d'écoute artificielles. Or, différents facteurs tels que l'ILD (Zhang et Mao, 2010), les différences interaurales de phase (Mulligan *et al.*, 1985), la corrélation interaurale (Edmonds et Culling, 2009) ou encore la répartition de l'énergie spectrale entre les deux oreilles (Zwicker et Zwicker, 1991) semblent avoir une influence sur la quantité de sommation de sonie binaurale. Ces différents indices binauraux sont fortement impliqués dans le processus de localisation des sources et la perception de notre environnement, ce qui suggère notamment un lien entre la position spatiale de la source et la sonie. Cela souligne finalement l'intérêt de considérer des situations d'écoutes plus réalistes, afin d'étudier l'influence de ces différents paramètres sur des plages de valeurs et de variation "naturelles". Pour s'approcher d'une situation réelle d'écoute et permettre un rendu spatial, il est donc nécessaire de prendre en compte la transformation acoustique qui opère entre le champ acoustique (sans obstacle) et chacune des oreilles de l'auditeur (combinaison des HRTFs (Head Related Transfer Functions) en champ libre et de l'acoustique du lieu). Ne pas prendre en compte ces transformations mène la plupart du temps à une localisation (plus ou moins latéralisée) des sources "à l'intérieur de la tête" de l'auditeur et peut donc réduire la validité écologique des résultats par rapport à une situation réelle avec des sources spatialisées.

1.3 Sonie binaurale d'un champ acoustique

Dans un champ acoustique réel, on peut considérer que l'écoute est toujours dichotique. En effet, nous utilisons nos deux oreilles, et les signaux à l'oreille gauche et à l'oreille droite diffèrent en fonction de la position de la source, de sa distance et de l'acoustique du lieu. L'étude de la sonie binaurale pour des champs acoustiques réels suggère donc une décomposition en deux parties : une partie acoustique visant à décrire la transformation entre le champ acoustique et les signaux reçus aux oreilles de l'auditeur, et une partie psychoacoustique visant à comprendre comment à partir de ces signaux reçus à chaque oreille un unique percept de sonie globale est créé. Dans les sections suivantes nous décrirons comment la variation des propriétés du champ acoustique (position de la source, nombre de sources, type de champ,...) influent sur la sonie binaurale et dans quelle mesure il est possible de relier directement la sonie binaurale aux variations induites dans les signaux de pression reçus aux oreilles de l'auditeur.

1.3.1 Sonie binaurale d'une source en champ libre

Le champ acoustique le plus simple pouvant être étudié est constitué d'une seule source en champ libre. Le champ libre n'est pas une situation d'écoute naturelle mais permet d'isoler uniquement le son direct de la source en s'affranchissant des réflexions. Dans ces conditions les signaux reçus aux oreilles de l'auditeur varient suivant l'angle d'incidence de la source. Ces variations dépendent directement du filtrage physique opéré par le buste, la tête et les pavillons de l'auditeur, filtrage caractérisé pour chaque oreille et chaque direction de l'espace par une fonction de transfert HRTF. L'HRTF est définie plus précisément comme la fonction de transfert entre la pression à l'oreille de l'auditeur et la pression mesurée au centre de la tête de l'auditeur en son absence (Møller, 1992). Ainsi, pour une source située dans le plan sagittal ($\Phi = 0^\circ$ ou 180°), les HRTFs à l'oreille gauche et à l'oreille droite sont similaires (due à la symétrie du corps), la situation d'écoute est donc approximativement dichotique. En revanche, lorsque la source sort du plan sagittal, des différences sont introduites entre les HRTFs gauche et droite. Les deux principaux indices binauraux extraits des HRTFs pour caractériser ces différences sont l'ITD (différence interaurale de temps) et ILD (différence interaurale de niveau). La dépendance spatiale des HRTFs et des indices binauraux est souvent considérée comme la clé pour la localisation des sources (e.g. Blauert (1997)). Cependant, peu d'études se sont intéressées directement à l'implication de cette dépendance en terme de sonie binaurale.

Robinson et Whittle (1960) furent les premiers à avoir mesuré l'effet de l'angle d'incidence d'une source sur la sonie. Les stimuli étaient des bandes de bruit tiers d'octave centrées à 1.6, 2.5, 4, 6.4, 8 ou 10 kHz. Les sons étaient restitués soit dans le plan horizontal ($\Theta = 0^\circ$), soit dans le plan sagittal ($\Phi = 0^\circ$ ou 180°), soit dans le plan transverse ($\Phi = \pm 90^\circ$), à l'aide d'une série de haut-parleurs espacés régulièrement par pas de 30° . L'auditeur avait pour tâche d'ajuster le niveau d'un son de comparaison suivant une certaine direction pour qu'il ait la même sonie que ce même son suivant une incidence de référence. L'égalisation en sonie suivant les différentes directions a permis aux auteurs d'établir des courbes de niveau d'isonomie en fonction de l'angle d'incidence et de la fréquence. Ces résultats ont ainsi mis en évidence un effet de l'angle d'incidence sur la sonie globale, effet plus prononcé pour les fréquences hautes. De plus, les niveaux de pression aux oreilles de chaque auditeur ont été mesurés dans les différentes configurations de l'expérience. Des similarités importantes ont

été constatées entre l'évolution de la sonie en fonction de l'angle d'incidence et celle des niveaux de pression aux oreilles. Un modèle de sommation de sonie binaurale a alors été proposé par Robinson et Whittle (1960), suggérant de calculer un niveau monaural équivalent L_{mon} (produisant la même sonie que le son binaural) en combinant les niveaux de pression à l'oreille gauche (L_G) et à l'oreille droite (L_D) :

$$L_{mon} = g \times \log_2(2^{L_G/g} + 2^{L_D/g}) \quad (1.6)$$

avec $g = 6 \text{ dB}$, correspondant au gain de sommation binaurale permettant d'approcher au mieux les données subjectives.

Une expérience similaire a été menée plus récemment par Sivonen et Ellermeier (2006) pour des bandes de bruit tiers d'octave également, mais centrées à 0.4, 1 et 5 kHz. Les azimuts 0, 30, 60, 90, 135 et 180° ont été testés dans le plan horizontal, ainsi que les élévations 45 et 90° dans le plan sagittal. Les niveaux d'égalisation en sonie ont été mesurés à l'aide d'une procédure 2I-2AFC, et le son de référence correspondait toujours à l'incidence frontale. Le test a été réalisé pour deux niveaux de référence différents (45 et 65 dB SPL). Les résultats obtenus à 65 dB SPL, moyennés sur les 8 auditeurs, sont présentés dans la figure 1.6. A nouveau, une variation importante de la sonie avec l'angle d'azimut a été observée, l'amplitude de variation augmentant avec la fréquence. La sensibilité directionnelle en sonie (DLS), définie comme l'opposé de la différence entre le niveau d'égalisation et le niveau de référence, pouvait varier de 8 dB entre deux angles d'incidence à 5 kHz. Le niveau global avait cependant une influence faible sur la DLS (ce résultat n'apparaît pas sur la figure).

Le modèle de sommation binaurale suggéré par Robinson et Whittle (1960) (eq. 1.6) a été ajusté sur les données de Sivonen et Ellermeier (2006). La valeur du gain binaural $g = 3 \text{ dB}$ a finalement été retenue par les auteurs. Les carrés sur la figure 1.6 représentent le niveau diotique équivalent calculé suivant cette règle ($L_{diot} = L_{mon} - g$) et montrent la performance du modèle pour approcher les données moyennes.

L'expérience a été répétée par les auteurs en simulant les sources au casque, via le processus de synthèse binaurale individuelle (Sivonen *et al.*, 2005) ou générique (Sivonen et Ellermeier, 2008). Elle a montré d'une manière générale que le recours à des sources virtuelles n'avait pas d'influence sur les égalisations en sonie individuelles, ni sur la valeur du gain binaural optimal pour approcher les données moyennes. Une variabilité interindividuelle du même ordre de grandeur a été rapportée par les auteurs.

Les résultats sur la sonie directionnelle en champ libre ont été étendus à des bruits large bande (bruit rose) dans le plan horizontal (Sivonen, 2007). Bien que de manière moins importante, la sonie variait également en fonction de l'angle d'azimut (plage de variation de la DLS inférieure à 4 dB, contre 8 dB pour un tiers d'octave centré à 5 kHz). Le modèle initialement proposé par Robinson et Whittle (1960) pour des bruits à bande étroite a été généralisé pour des sons complexes (Sivonen et Ellermeier, 2008). Sivonen (2007) a montré que ce modèle permettait alors de prédire la sensibilité directionnelle en sonie de manière précise pour des bruits large-bande également (toujours avec $g = 3 \text{ dB}$). Cabrera *et al.* (2008) ont également mesuré la sonie de sources spatialisées pour des bruits d'une octave centrés à 125, 250, 500, 1000, 2000, 4000 ou 8000 Hz et pour des azimuts de $\pm 30^\circ$ ou $\pm 110^\circ$ (en configuration de restitution 5.1). La sonie variait avec la fréquence principalement pour les angles $\pm 110^\circ$ et le modèle à 3 dB de Sivonen et Ellermeier (2008) permettait également d'approcher

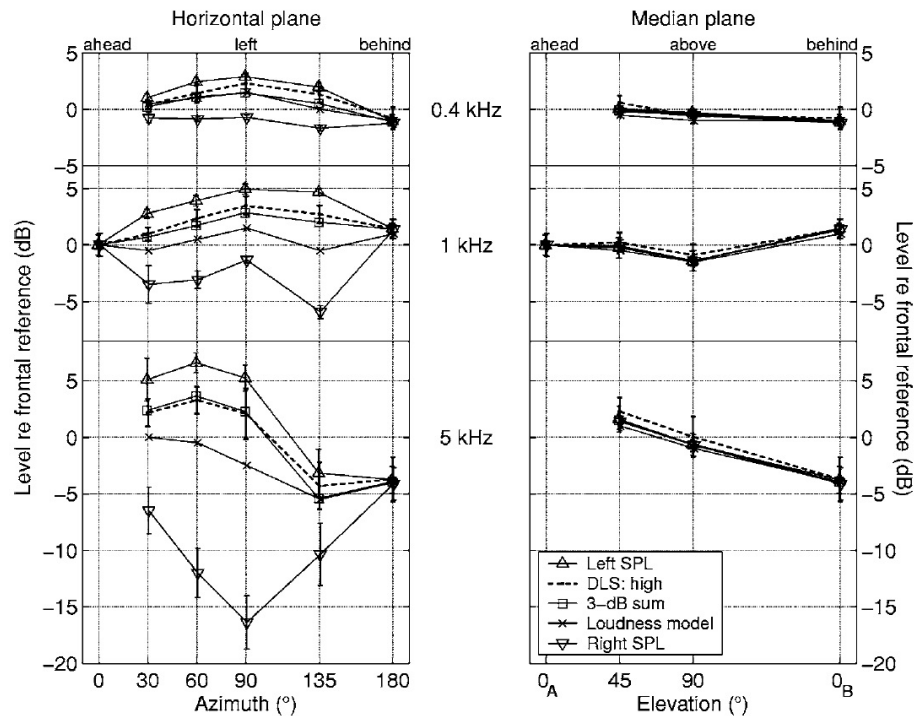


FIGURE 1.6: Fig. 8 de Sivonen et Ellermeier (2006). Niveaux de pression moyens relatifs (par rapport à 65 dB SPL) à l'oreille gauche et à l'oreille droite (triangles), loi de sommation à 3 dB (carrés), sommation de sonie en sones (croix) et DLS moyennes mesurées pour le niveau global de 65 dB SPL (pointillés).

correctement les données subjectives.

Dans l'ensemble de ces études, une variabilité interindividuelle importante a été mesurée dans les sensibilités directionnelles en sonie individuelles. Une variabilité importante persiste dans les valeurs individuelles du gain binaural estimé, malgré la prise en compte individualisée des HRTFs. Une hypothèse avancée dans Sivonen et Ellermeier (2006) est basée sur le phénomène de "constance perceptuelle" en sonie mis en évidence par Zahorik et Wightman (2001) (cf. § 1.3.4.1). Selon cette hypothèse, les auditeurs auraient une tendance plus ou moins grande à juger la sonie constante quelque soit l'angle d'incidence, malgré les différences interaurales que cela implique. Cependant, aucune explication justifiée n'a pour le moment permis de comprendre l'origine de ces différences entre les auditeurs. L'influence de la variabilité des fonctions de sonie monaurales individuelles sur la variabilité interindividuelle mesurée dans les DLS individuelles fait l'objet du chapitre 3.

L'étude de la sonie d'une source de bruit spatialisée en champ libre soutient finalement la théorie d'une sommation de sonie moins que parfaite. On notera que le modèle de sommation binaurale proposé (Sivonen et Ellermeier, 2008) est basé uniquement sur les différences interaurales de niveau par bande de fréquence, et ne tient donc pas compte des différences interaurales de temps ni de l'effet de la corrélation interaurale susceptible d'intervenir pour des signaux plus complexes (Edmonds et Culling, 2009).

1.3.2 Sonie binaurale de plusieurs sources en champ libre

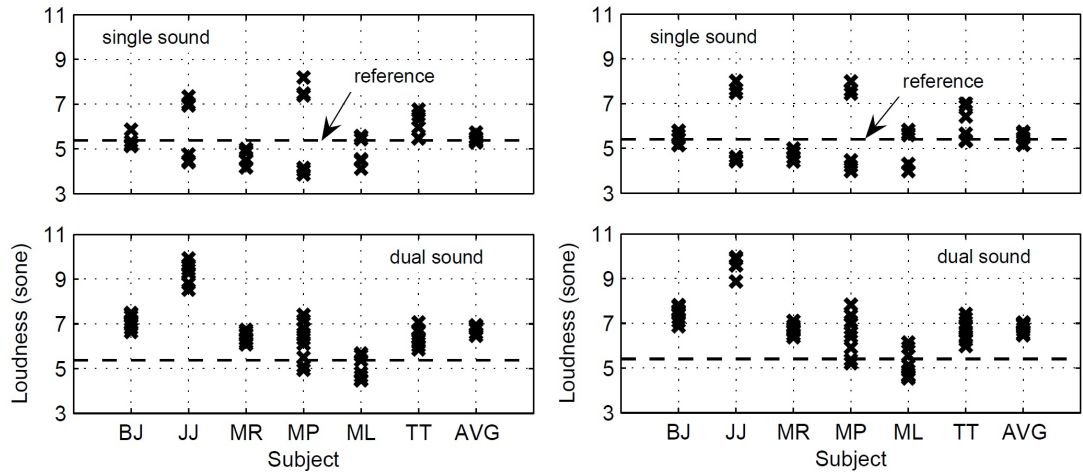
Très peu d'études se sont intéressées à la sonie des champs acoustiques en présence de plusieurs sources spatialisées simultanées.

Scharf (1974) a montré qu'en présence de deux sources, les propriétés de la sommation de sonie spectrale (cf. § 1.1.2.3) étaient conservées et indépendantes de la position réelle ou apparente des deux sources. En effet, lorsque deux sons purs de fréquences différentes provenaient de deux sources réelles séparées de 90° , le niveau de sonie augmentait de 3 phones lorsque les deux fréquences étaient contenues à l'intérieur de la même bande critique, puis la sonie augmentait avec la séparation en fréquence au-delà de la bande critique, de manière similaire au cas où les deux sinusoïdes sont présentées au casque à la même oreille. Il en était de même lorsqu'une source était retardée (jusqu'à 20 ms) par rapport à l'autre et que les deux sons semblaient provenir d'un seul des deux haut-parleurs (effet Haas). D'après Scharf (1974), ni la multiplicité des sources, ni la localisation apparente des sources (sans variation des différences interaurales) n'a d'impact sur la sommation de sonie spectrale. Cela suggère donc que la sonie d'un son complexe pourrait être estimée à partir des signaux reçus aux oreilles de l'auditeur, indépendamment de la direction de chaque composante.

L'effet de la multiplicité et de la distribution spatiale (réelle) des composantes sur la sonie globale n'avait pas été testé directement dans Scharf (1974), mais a fait l'objet d'une partie des travaux de Song (2007). Cette étude avait pour objectif de tester l'effet de la distribution spatiale de deux sources tonales simultanées (bandes de bruit tiers d'octave centrées à 1 et 3.15 kHz) sur la sonie globale. Les sources étaient soit focalisées en incidence frontale, soit distribuées suivant des angles d'azimut de $\pm 10^\circ$ ou $\pm 30^\circ$. La sonie était toujours évaluée à l'aide d'une procédure adaptative 2I-2AFC consistant à égaliser la sonie de la composante à 3.15 kHz pour que le champ acoustique complexe produise la même sonie qu'une référence d'incidence frontale centrée à 2 kHz. Les composantes tonales (1 kHz ou 3.15 kHz) spatialisées suivant un azimut non-nul étaient préalablement égalisées en sonie par rapport à la référence frontale (2 kHz) pour s'affranchir de la sensibilité directionnelle en sonie propre à chaque composante (Sivonen et Ellermeier, 2006). Finalement, la sonie dans le cas où les sources étaient distribuées était comparée à la sonie dans le cas de sources focalisées, menant à conclure que la distribution des sources tonales dans le plan horizontal n'avait pas d'influence sur la sonie globale (pour des angles de 10° et 30°).

La capacité des modèles de sonie à prédire ces résultats a également été mise à l'épreuve. Les champs acoustiques générés par la référence (2 kHz) et par les différentes combinaisons de sources du protocole ont été mesurés de deux façons : d'une part à l'aide d'un microphone placé au centre de la tête de l'auditeur en son absence, puis à l'aide d'une tête artificielle. Les figures 1.7a et 1.7b correspondent aux figures 6 et 7 de Song (2007) (p. 50). Elles représentent les prédictions de deux modèles (en tenant compte des égalisations individuelles ou moyennes en sonie) : le modèle diotique ISO-532B (1975) appliqué à la mesure du microphone seul (figure 1.7a) et le modèle binaural de Sivonen et Ellermeier (2008) appliqué aux mesures sur la tête artificielle (figure 1.7b).

Les graphiques du haut correspondent aux égalisations préliminaires en sonie pour chaque composante tonale indépendamment ("single sound"), tandis que les graphiques du bas correspondent aux égalisations en présence des deux sources tonales simultanées ("dual sound"). Les estimations sur les données moyennes ont montré d'une manière générale que les modèles fonctionnaient pour prédire la sonie d'une seule composante tonale (en haut). En revanche, tous les modèles tendaient



(a) Prédiction de la sonie par l'ISO-532B (1975) (présentation diotique).

(b) Prédiction de la sonie par le modèle de Sivonen et Ellermeier (2008) (présentation dichotique).

FIGURE 1.7: Figures 6 et 7 de Song (2007). Prédiction pour les champs acoustiques produits par chaque auditeur en égalisant la sonie : d'une composante tonale (en haut) ou de deux composantes tonales (en bas) à la sonie de la référence frontale (pointillés). Six (en haut) ou neuf (en bas) valeurs de sonie correspondant aux différentes conditions expérimentales sont représentées. Les haut-parleurs étaient placés à $\pm 10^\circ$ d'azimut et le niveau de la référence (2 kHz) était fixé à 60 dB SPL.

à surestimer légèrement la sonie en présence de deux composantes tonales (en bas), même dans le cas où les deux sources étaient focalisées en incidence frontale. L'accord entre les deux modèles s'explique par le faible angle d'incidence des sources ($\pm 10^\circ$) impliquant des faibles différences interaurales. Ces résultats soulignent néanmoins l'intérêt d'étudier la sonie binaurale en présence de plusieurs sources réelles, dans le but d'améliorer la prédiction des modèles dans ce sens.

L'expérience présentée dans le chapitre 4 constitue une extension des résultats de Song (2007) pour des grands angles d'incidence (60° et 90°). Dans notre cas, la sonie des sources distribuées a été directement comparée à la sonie des sources focalisées mais suivant le même angle d'incidence (contrairement à l'étude de Song (2007) qui se rapporte à l'incidence frontale). La précision des modèles pour prédire la sonie d'un champ acoustique bitonal y est à nouveau évaluée.

1.3.3 Sonie binaurale en-dehors du champ libre

Le champ libre permet d'isoler le son direct d'une source, en s'affranchissant de l'acoustique de l'environnement dans lequel s'inscrit cette source. Cependant, les conditions du champ libre n'existent pas dans la réalité où les réflexions des ondes acoustiques sont omniprésentes. Il est donc nécessaire de s'intéresser à l'effet de la réverbération sur la sonie afin d'améliorer la validité des modèles. Le cas idéal de champ réverbéré est le champ diffus : champ acoustique homogène et isotrope. Autrement dit, dans un champ diffus les ondes sonores arrivent aux oreilles de l'auditeur avec la même énergie dans toutes les directions, contrairement au champ libre où les ondes sonores proviennent uniquement de la direction de la source.

L'effet de la réverbération dans le cas du champ diffus peut être considéré à deux niveaux : d'un point de vue physique en comparant les signaux reçus aux oreilles de l'auditeur en champ libre ou en champ diffus et d'un point de vue psychophysique en mesurant la différence de sonie correspondante.

Ainsi, la courbe de niveau d'isotonie à 70 phones en champ diffus a été mesurée par Stevens (1956) pour des bruits en bandes d'octave. La courbe obtenue différait des courbes d'isotonie mesurées en champ libre (e.g. Fletcher et Munson (1933)), notamment pour les fréquences hautes. Une étude similaire menée par Robinson *et al.* (1961) pour des bandes de bruit en tiers d'octave visait à comparer directement la sonie entre champ libre et champ diffus, comparaison rendue possible par la mise en place d'un champ diffus artificiel (20 haut-parleurs décorrélés). La différence de sonie entre champ diffus et champ libre estimée par Robinson *et al.* (1961) en associant leurs résultats aux données de quelques études antérieures (e.g. Kuhl et Westphal (1959)) est présentée dans la figure 1.8 issue de Sivonen et Ellermeier (2011).

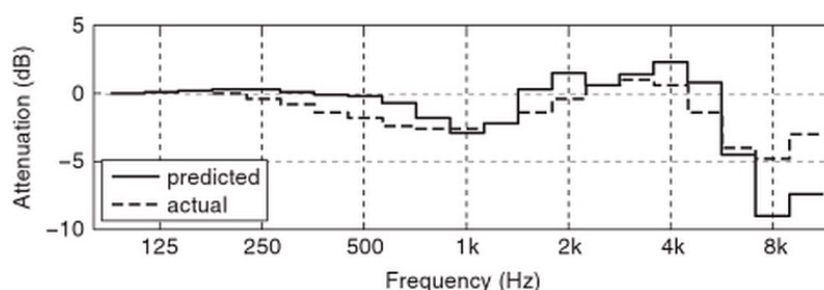


FIGURE 1.8: Fig. 7.5 de Sivonen et Ellermeier (2011). Atténuation pour un son en champ diffus au point d'égalité subjective en sonie avec ce même son en champ libre. Les données "prédites" (courbe continue) sont basées sur les mesures physiques d'une tête artificielle (ISO11904-2, 2004). Les données "réelles" (courbe pointillée) correspondent à des mesures psychophysiques de sonie (Robinson *et al.*, 1961).

La figure représente l'atténuation par bande de tiers d'octave à appliquer en champ diffus pour obtenir la même sonie qu'en champ libre (incidence frontale). Les résultats de Robinson *et al.* (1961) (courbe en pointillé) sont comparés à la différence de niveau de pression aux oreilles correspondante (valeurs dérivées d'HRTFs mesurées sur une tête artificielle (ISO11904-2, 2004)). La comparaison des deux courbes montre que la différence de sonie (mesure subjective) coïncide avec la différence de niveaux aux oreilles (mesure physique). Cela suggère donc que les différents processus d'intégration et de sommation de sonie binaurale à l'origine du percept de sonie globale sont indépendants du champ acoustique. Autrement dit, l'information sur la nature du champ acoustique est principalement comprise dans les niveaux de pressions reçus aux oreilles de l'auditeur et n'implique aucune modification dans le calcul de la sonie à partir des signaux mesurés à chaque oreille.

Cette théorie a été confortée pour des sources de bruit (bandes de tiers d'octave centrées à 1 et 5 kHz) spatialisées dans le plan horizontal (Sivonen, 2007). La salle simulée au casque avait un temps de réverbération de 0.44 s à 1 kHz et de 0.62 s à 5 kHz. Malgré les différences de niveaux de pression aux oreilles par rapport aux mêmes sources spatialisées en champ libre (Sivonen et Ellermeier, 2008), la règle de sommation binaurale restait inchangée (gain de sommation binaurale $g = 3 \text{ dB}$).

Actuellement, le seuil d'audition de référence en champ diffus est standardisé dans la norme

ISO389-7 (2005). De plus, une égalisation pour le champ diffus est disponible dans les modèles ISO-532B (1975) et ANSI-S3.4 (2007). Elle correspond à la fonction de transfert moyenne du champ diffus (sans obstacle) au tympan, pour une personne adulte.

Cependant, bien que l'effet de la diffusion du champ sur les ILDs semble en grande partie expliquer les différences de sonie, la diffusion du champ est également directement liée à la corrélation interaurale, pouvant influencer la sonie (Edmonds et Culling, 2009). Miranda et Cabrera (2008) ont évalué l'effet de la diffusion du champ acoustique sur la sonie pour des sources spatialisées à l'aide de hauts-parleurs (suivant des angles de 0° , $\pm 45^\circ$, ou $\pm 90^\circ$). Les stimuli étaient soit un bruit rose, soit une des trois sous-bandes : BF (octaves centrées à 63, 125 et 250 Hz), MF (octaves centrées à 500, 1000 et 2000 Hz) ou HF (octaves centrées à 4000, 8000 et 16000 Hz). Le champ acoustique était soit direct (un seul HP), soit diffus (bruits indépendants joués au même niveau par les 8 HPs du dispositif), soit composé du son direct (sur un HP) restitué 6 ou 12 dB plus fort que tous les autres bruits restitués sur les HPs restants constituant le champ diffus. Les auditeurs devaient ajuster le niveau d'un stimulus de comparaison pour qu'il ait la même sonie que le son direct d'incidence frontale (de même contenu spectral). D'après la figure 5 de Miranda et Cabrera (2008) présentant la sensibilité en sonie en fonction de la corrélation interaurale (comprise entre 0 et 1), il semblerait que d'une manière générale, la sonie augmente lorsque la corrélation interaurale diminue. De plus, bien qu'aucun effet simple de la diffusion sur la sonie n'ait été relevé par les auteurs, il y avait une interaction forte entre la diffusion et la fréquence. D'après la figure, cela se traduisait par un effet plus important de la diffusion sur la sonie pour les bandes de bruit BF et MF. Ces résultats rejoindraient alors les conclusions de Edmonds et Culling (2009) sur l'augmentation de la sonie lorsque la corrélation interaurale est nulle et que la majorité de l'énergie du stimulus est contenue en basse fréquence. Cela suscite donc l'intérêt d'étudier l'influence du champ acoustique sur la corrélation interaurale pouvant dans certaines conditions (a priori liées au spectre) avoir une influence sur la sonie.

1.3.4 Lorsque la sonie ne dépend plus uniquement des signaux reçus aux oreilles

Bien que la sonie puisse être prédite en grande partie à partir des signaux acoustiques reçus aux oreilles de l'auditeur, d'autres facteurs comme le contexte (interaction d'autres modalités sensorielles (vue, toucher,...), sources concurrentes,...), l'auditeur (état physique et psychologique (fatigue, pertes auditives,...), mémoire, connaissances, culture,...), ou l'analyse de scène auditive peuvent jouer un rôle parfois non négligeable. L'impact de ces autres facteurs sur la sonie est illustré à travers trois exemples dont l'influence potentielle mérite qu'on s'y attarde.

1.3.4.1 Constance perceptuelle de la sonie (*Loudness constancy*)

La constance perceptuelle est une notion de psychologie décrivant la capacité à reconnaître l'invariance d'un objet, indépendamment du contexte (position, distance, orientation,...) entraînant pourtant des modifications importantes de l'information du point de vue des différents récepteurs sensoriels (taille, couleur, spectre, niveau sonore,...). Un exemple classique de la perception visuelle décrit le fait que la taille d'une personne (dans un contexte) est perçue de la même manière quelle que soit la distance qui sépare l'observateur de la personne en question (et donc quelle que soit la taille de l'image correspondante sur la rétine de l'observateur). Or, il se trouve que l'analogie existe en

perception sonore.

En effet, Zahorik et Wightman (2001) ont montré que le jugement de sonie ne variait pas en fonction de la distance de la source (comprise entre 0.3 et 13.8 m), lorsqu'il était demandé à l'auditeur de porter le jugement sur la puissance de la source (consigne "objective"). Cette invariance a été mesurée malgré une variation des niveaux aux oreilles de l'auditeur de l'ordre de 20 dB entre les distances extrêmes. Les stimuli étaient des sources virtuelles créées en convoluant des bruits gaussiens de 50 ms par des réponses impulsionnelles binaurales de salle (individuelles) mesurées dans un auditorium ($TR_{60} \simeq 0.7$ s) pour chaque distance. Selon Zahorik et Wightman (2001), l'invariance en sonie avec la distance serait expliquée par le fait que les auditeurs basaient leurs jugements sur la quantité d'énergie réverbérée qui est relativement constante avec la distance dans les salles, et proportionnelle à la puissance de la source. En revanche, en l'absence de contexte (en champ libre), la constance perceptuelle en sonie n'est plus observée. Dans ce cas, l'auditeur base nécessairement son jugement de sonie sur l'énergie du son direct qui décroît avec la distance de la source (loi du carré inverse), d'où une décroissance de la sonie en conséquence (Stevens et Guirao (1962), Petersen (1990)). Une théorie intermédiaire plus complexe est que l'auditeur utiliserait le rapport entre l'énergie directe (qui varie avec la distance) et l'énergie réverbérée (approximativement constante) pour évaluer la distance de la source, puis juger la sonie en conséquence. Dans le cas où la consigne est uniquement de juger la sonie indépendamment de la source (consigne "subjective"), dans un contexte offrant la possibilité à l'auditeur d'évaluer la distance de la source et donc sa puissance, le jugement de sonie pourrait alors être biaisé par l'idée que se fait l'auditeur de la puissance de la source.

La connaissance de la source augmente la précision dans l'évaluation de sa distance. La constance perceptuelle en sonie est ainsi favorisée par la présence visuelle de la source (Macpherson, 1995). De la même manière, des sources réelles comme de la parole ou de la musique permettent une meilleure constance perceptuelle en sonie que les sons purs ou les bruits.

On constatera que dans l'étude de Sivonen (2007), malgré la spatialisation des sources de bruit à distance constante dans une acoustique réverbérante, la sonie variait avec l'angle d'incidence de la source (et donc avec la variation des niveaux de pression aux oreilles qui en découlent). Dans ce cas, l'effet de constance perceptuelle n'est pas apparu clairement dans les résultats. La consigne consistait cependant à choisir le son le plus fort dans chaque paire de sons (consigne "subjective") et ne précisait donc pas de porter le jugement sur la puissance de la source (consigne "objective"), contrairement à Zahorik et Wightman (2001).

1.3.4.2 Interactions multisensorielles

Dans un environnement réel, l'information est captée par nos différents sens puis intégrée par le système nerveux pour permettre une représentation cohérente des différents objets. L'intégration est multisensorielle dans le sens où les différentes modalités sensorielles peuvent interagir entre elles, ouvrant la possibilité que chacune puisse ainsi modifier le processus de toutes les autres. Dans ce sens, plusieurs types d'interactions multisensorielles pouvant avoir une influence sur la sonie ont été mis en avant.

L'interaction audio-visuelle est vraisemblablement la plus étudiée. Par exemple, Odgaard *et al.* (2004) ont mis en évidence l'augmentation (très claire) de la sonie pour des bruits de 40 ms à faible niveau (entre 39 et 43 dB(A)) en présence de flashes lumineux synchrones de même durée. L'interac-

tion a également été mise en évidence pour des images, lorsque ces dernières ont un sens en rapport avec les sons présentés. Un bruit blanc est perçu moins fort si il est associé à une image de cascade (Suzuki *et al.*, 2000), de la même manière qu'un train qui passe est perçu moins fort si l'image accompagne le son (Fastl, 2004). De plus, plus l'interaction est réaliste (image filmée en rapport avec le son, mouvement de la caméra, dispositif de restitution,...), plus elle semble avoir un impact important sur la réduction de la sonie (Fastl, 2004). Une expérience antérieure sur l'atténuation des barrières acoustiques montrait déjà l'impact important du facteur visuel sur la sonie dans une situation réaliste (Aylor et Marks, 1976). Dans la situation plus classique d'une voix parlée, Epstein et Florentine ont montré que la sommation de sonie binaurale était plus faible lorsque le locuteur était présent visuellement (Epstein et Florentine (2009), Epstein et Florentine (2012)).

La couleur peut également avoir une influence : un train rouge est généralement perçu plus fort qu'un train vert (Patsouras *et al.*, 2002) (effet observé pour des auditeurs allemands et japonais (Rader *et al.*, 2004), mais pas pour des auditeurs français (Parizet et Koehl, 2011)).

L'existence d'une interaction audio-tactile a également été mise en évidence. Ainsi, selon Gilmeister et Eimer (2007), l'ajout d'une stimulation vibro-tactile synchrone au stimulus sonore (à faible niveau) a pour effet d'augmenter la sonie. En revanche, une stimulation tactile intense a pour effet d'augmenter légèrement le seuil d'audition, phénomène interprété comme du masquage intermodal (Gescheider et Niblette, 1967). Une revue plus complète sur l'influence des interactions entre les différentes modalités sensorielles sur la sonie a été établi par Fastl et Florentine (2011).

1.3.4.3 Sonie et Analyse de Scène Auditive (ASA)

L'analyse de scène auditive désigne le processus par lequel, à partir de la superposition complexe des sons au niveau des récepteurs sensoriels, le système auditif organise l'information (par groupement ou ségrégation) en éléments perceptifs ayant un sens (Bregman, 1990). Il est donc vraisemblable que la sonie globale d'un son (normalement obtenue par la sommation des sonies spécifiques) pourra être affectée par la manière dont les éléments qui le constituent seront groupés entre eux. Cependant, très peu d'études se sont intéressées à la manière dont les processus de groupement pouvaient influencer la sonie.

McAdams *et al.* (1998) ont mené des expériences dont les stimuli étaient des séquences alternées de deux sons brefs ayant le même contenu spectral mais des niveaux différents ($L_1 < L_2$). Lorsque la présentation était diotique et que l'alternance des deux sons se faisait sans silence intermédiaire, la séquence était perçue comme la somme d'un son continu et d'un son intermittent (illusion de continuité auditive). Lorsqu'il était demandé à l'auditeur d'ajuster en sonie le niveau d'un son de comparaison intermittent (L_{aj}) à la composante perçue comme intermittente (L_2) dans le son de référence soumis à l'illusion perceptive, le niveau d'ajustement pouvait être atténué jusqu'à 12 dB par rapport au niveau physique réel de la composante ($L_{aj} < L_2$). En effet, les parties de niveau L_2 dans le son de référence étaient interprétées comme la somme de la continuité du son précédent de niveau L_1 et d'un son superposé de niveau $L_2 - L_1$, chacun ayant sa propre sonie. Ces résultats fournissent donc un exemple concluant de la manière dont l'organisation perceptive d'une séquence peut modifier la sonie.

Pour des sons éloignés en fréquence, l'intégration ou la ségrégation des composantes peut également avoir une influence sur la sonie. En effet, Grimault *et al.* (2007) ont montré que lorsque les

composantes d'un son complexe (1 kHz + 2 kHz) étaient ségréguées, une des composantes pouvait donner lieu à une sonie en particulier. En revanche lorsque les deux composantes étaient intégrées, chaque composante contribuait alors à un percept de sonie globale. La sonie globale mesurée était cependant inférieure aux prédictions de la sommation de sonie sur les différentes bandes critiques (Fletcher et Munson, 1933), ce qui reflétait peut-être des organisations perceptives différentes selon les auditeurs (intégration plus ou moins claire des composantes).

Ces deux expériences montrent donc que la perception de la sonie peut être fortement influencée par les mécanismes de l'analyse de scène auditive, à l'intérieur d'une même bande critique (McAdams *et al.*, 1998) ou bien entre différentes bandes critiques (Grimault *et al.*, 2007).

Conclusion L'ensemble de ces exemples souligne finalement la nécessité de mieux comprendre les influences (parfois importantes) de certains facteurs physiologiques, cognitifs ou contextuels qui n'interviennent pas de manière physique dans les signaux reçus aux oreilles de l'auditeur mais qui peuvent avoir un rôle important dans le calcul de la sonie. Négliger ces facteurs peut, dans de nombreux cas réalistes, mener à des erreurs de prédiction importantes.

1.4 Modèles de sonie binaurale

Cette partie vise dans un premier temps à rappeler quels sont les principaux modèles de sonie binaurale existants et à décrire brièvement leurs méthodes de calcul. Nous proposons ensuite de discuter des limitations de ces modèles à travers quelques exemples et simulations. Tous les modèles présentés ici se limitent à la prédiction de la sonie pour des sons stationnaires.

1.4.1 Présentation des différents modèles

La construction des modèles de sonie binaurale est généralement basée sur la combinaison entre une méthode de calcul de la sonie en présentation diotique et d'un modèle de sommation de sonie binaurale permettant de prendre en compte l'aspect dichotique.

1.4.1.1 Calcul de sonie en présentation diotique

Actuellement, deux principales méthodes de calcul ont été développées pour calculer la sonie des sons stationnaires : la méthode de Zwicker (1958) ayant fait l'objet de la norme internationale (ISO-532B, 1975) puis d'une norme allemande DIN-45631 (1990) et la méthode de Moore *et al.* (1997) ayant fait l'objet de la norme américaine ANSI-S3.4 (2007).

Nous nous intéresserons ici plus particulièrement aux deux modèles les plus utilisés : les modèles ISO-532B (1975) et ANSI-S3.4 (2007). Ces deux modèles permettent de calculer la sonie d'une source d'incidence frontale en champ libre (situation diotique), à partir du signal acoustique (en Pa) enregistré au centre de la tête de l'auditeur en son absence. Leurs structures de base sont similaires, consistant à calculer d'abord la transformation du signal acoustique en niveau d'excitation du système périphérique (étapes 1 à 3), puis la transformation de l'excitation en sonie spécifique (étape 4), puis enfin la sonie globale en intégrant la sonie spécifique sur l'ensemble des filtres auditifs (étape 5). La figure 1.9 détaille les différentes étapes communes aux deux modèles (figure tirée de Molla *et al.* (2010)).

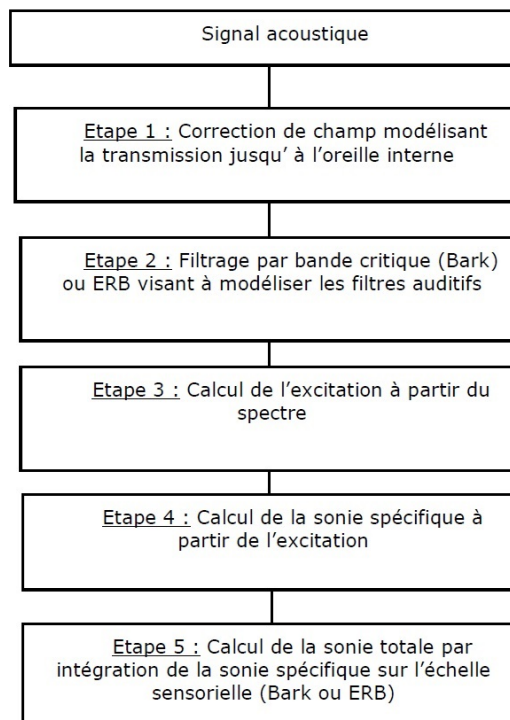


FIGURE 1.9: Fig. 1 de Molla *et al.* (2010). Étapes de calcul de la sonie pour des sons stationnaires.

Étape 1 :

La première étape consiste à prendre en compte la transformation physique du signal acoustique, depuis le champ libre (sans obstacle) jusqu'à l'oreille interne. Cette transformation correspond donc théoriquement aux fonctions de transfert du champ libre au tympan (buste, tête, oreille externe) et de l'oreille moyenne.

Dans l'ISO-532B (1975), la réponse en fréquence correspondante est modélisée par un unique filtre linéaire, dont l'atténuation correspondante est calculée pour chaque bande critique (cf. Etapes 2 et 3). Une correction égale à zéro en dessous de 2 kHz puis égale à l'inverse du seuil d'audition au-delà de 2 kHz est ainsi appliquée. L'augmentation du seuil d'audition en basses fréquences est supposée provenir du bruit interne de l'oreille (plus élevé en basses-fréquences), effet pris en compte dans l'étape 4.

Dans l'ANSI-S3.4 (2007), la réponse en fréquence du champ libre au tympan (HRTF en incidence frontale) est corrigée indépendamment de la réponse en fréquence de l'oreille moyenne (filtre basé sur des mesures de l'oreille moyenne en-dessous de 500 Hz, égal au seuil d'audition inversé au tympan au-delà de 500 Hz).

Les deux modèles proposent également une égalisation correspondant au champ diffus. Elle consiste en une correction supplémentaire dans le cas de l'ISO-532B (1975) ou à utiliser une HRTF mesurée en champ diffus pour l'ANSI-S3.4 (2007).

Étapes 2 et 3 :

Dans l'oreille interne, l'énergie transmise à travers la fenêtre ovale est répartie le long de la membrane basilaire, dont le fonctionnement est analogue à l'application d'une série de filtres fréquentiels. La modélisation de ces filtres physiologiques diffère selon les modèles.

La première représentation introduite par Zwicker (1961) est basée sur l'échelle des Barks définissant 24 bandes critiques juxtaposées représentatives de l'organisation tonotopique de la cochlée. Dans chaque bande critique, l'excitation est dérivée des diagrammes de masquage d'un son pur par un bruit et varie donc avec le niveau et la fréquence. C'est selon ce modèle que le diagramme d'excitation est calculé dans l'ISO-532B (1975), correspondant à l'énergie totale contenue dans chacune des 24 bandes critiques (en prenant en compte les effets de masquage fréquentiel).

L'ANSI-S3.4 (2007) utilise également une série de filtres passe-bande représentant la sélectivité fréquentielle à différentes positions sur la membrane basilaire. La largeur de bande de ces filtres est définie par des ERBs (pour Equivalent Rectangular Bandwidth), notion proche des bandes critiques de Zwicker, introduite et définie par Glasberg et Moore (1990). La largeur de bande d'une ERB est définie à niveau moyen par la relation suivante : $ERB = 24.7(0.00437 \times f + 1)$ où f est la fréquence centrale en Hz. La largeur de bande et la forme des filtres dépendent également du niveau. Le niveau d'excitation en sortie d'un filtre auditif (ERB_N) est calculé en sommant la puissance en sortie du filtre en réponse à l'ensemble du spectre du signal d'entrée. Le diagramme d'excitation correspond au niveau de sortie des filtres auditifs (en dB), en fonction de leurs fréquences centrales (résolution de $0.1ERB_N$).

Étapes 4 et 5 :

Le diagramme d'excitation est ensuite converti en sonie spécifique N' , représentant la contribution de chaque bande critique ou ERB à la sonie globale.

Dans le modèle ISO-532B (1975), la relation entre la sonie spécifique et le diagramme d'excitation est basée sur la loi de Stevens et définie par une équation de la forme :

$$N' = C \left(\frac{E_{TH}}{E_0} \right)^\alpha \left[\left(0.5 + 0.5 \frac{E}{E_{TH}} \right)^\alpha - 1 \right] \quad (1.7)$$

où N' est la sonie spécifique (en sone/Bark), E est l'excitation à la fréquence concernée, E_{TH} est l'excitation au seuil d'audition produite par le bruit interne de l'oreille en fonction de la fréquence, E_0 est l'excitation de référence (pour $I_0 = 10^{-12} \text{ W/m}^2$), α correspond à l'exposant de la fonction de sonie (fixé à 0.23) et C est une constante. La sonie globale est ensuite calculée en sommant la sonie spécifique sur les 24 bandes critiques. C est ajustée pour que la sonie globale d'un son pur de 1 kHz à 40 dB SPL en champ libre soit égale à 1 sone.

Dans le modèle ANSI-S3.4 (2007), la sonie spécifique N' est également reliée au diagramme d'excitation en fonction de l'exposant des fonctions de sonie α et de l'excitation au seuil E_{TH} . Cette relation non linéaire est définie de la manière suivante pour $E_{TH} < E < 10^{10}$:

$$N' = C [(GE + A)^\alpha - A^\alpha] \quad (1.8)$$

où G représente le gain d'amplification de la cochlée à bas niveau. A est un paramètre valant $2.E_{TH}$ au-delà de 500 Hz. Comme pour le modèle de Zwicker, la sonie globale est calculée en intégrant la

sonie spécifique sur l'ensemble des ERBs. La sonie globale ainsi obtenue correspond en revanche à la sonie d'une seule oreille. Dans une situation diotique d'écoute, les sonies obtenues pour chaque oreille sont simplement sommées entre elles. La constante C est ajustée pour que la sonie globale (diotique) d'un son pur de 1 kHz à 40 dB SPL en champ libre soit égale à 1 sone.

1.4.1.2 Modélisation de la sommation de sonie binaurale

Dans la plupart des situations réelles, l'écoute est dichotique. Une mesure monophonique du champ acoustique ne suffit plus pour estimer correctement la sonie. Il est alors nécessaire de réaliser des mesures binaurales (à l'aide d'une tête artificielle par exemple) afin de prendre en compte la contribution des signaux de chaque oreille dans la création du percept de sonie globale. Plusieurs modèles de sommation de sonie binaurale ont alors été proposés, permettant de prendre en compte les différences interaurales de niveau dans le calcul de la sonie. Chaque modèle vise à définir une règle de sommation binaurale simple offrant une représentation la plus fidèle possible des données subjectives recueillies dans la littérature. En revanche, aucun modèle actuel ne permet de prendre en compte les différences interaurales de temps ou la corrélation interaurale, en partie à cause du nombre restreint de résultats existant actuellement sur le sujet.

Actuellement, les trois principaux modèles permettant de calculer la sonie binaurale en situation dichotique d'écoute à partir d'un enregistrement réalisé sur un mannequin acoustique sont : **a)** le modèle de sommation de sonie parfaite standardisé dans la norme ANSI-S3.4 (2007), d'après Moore *et al.* (1997) **b)** le même modèle, augmenté par la prise en compte des inhibitions interaurales (Moore et Glasberg, 2007) **c)** le modèle proposé par Sivonen et Ellermeier (2008). Les principales étapes de calcul de ces trois modèles sont schématisées dans la figure 1.10 et détaillées dans les paragraphes suivants.

a) Sommation de sonie parfaite :

La sommation de sonie parfaite est le modèle le plus simple, basé sur l'hypothèse que la sonie globale N peut être décrite comme la somme de la sonie pour l'oreille gauche (N_G) et de la sonie pour l'oreille droite (N_D) :

$$N = N_G + N_D \quad (1.9)$$

Cette hypothèse est supposée valide, que la présentation des sons soit diotique (e.g. Fletcher et Munson (1933), Hellman et Zwischlocki (1963)) ou dichotique (e.g. Marks (1978)). En situation diotique, la sonie binaurale est égale au double de la sonie monaurale (rapport de sonie binaural-à-monastral égal à 2). Comme mentionné précédemment, c'est le modèle adopté par la norme américaine ANSI-S3.4 (2007), dans le cas où les signaux sont mesurés à l'aide d'une tête artificielle (si le champ acoustique est mesuré à l'aide d'un seul microphone, la sonie binaurale est calculée en doublant la sonie monaurale).

b) Modèle de Moore et Glasberg (2007) :

Bien que la sommation de sonie parfaite ait été longtemps soutenue, des études plus récentes se sont positionnées en faveur d'une sommation de sonie moins que parfaite (e.g. Scharf (1969), Algorn *et al.* (1989b), Zwicker et Zwicker (1991), voir partie 1.2.1). Ainsi, Moore et Glasberg (2007) ont

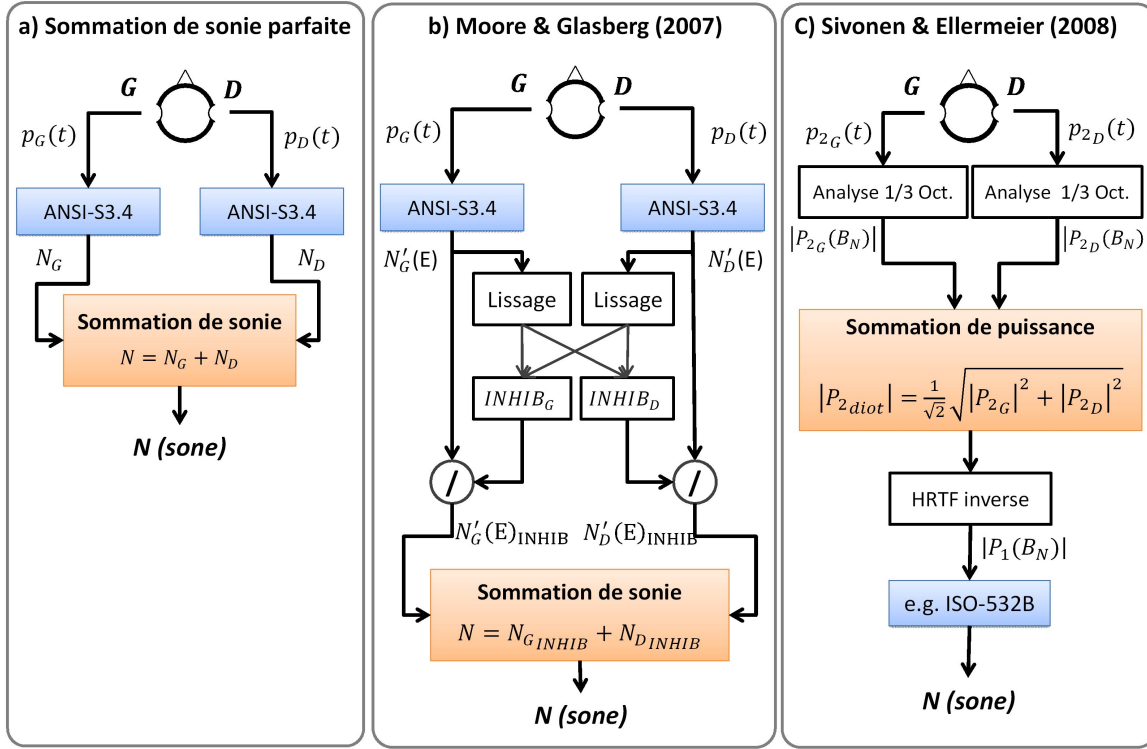


FIGURE 1.10: Représentation de la structure générale des trois modèles de sonie binaurale présentés.

proposé des modifications à apporter au modèle de Moore *et al.* (1997) pour décrire la sommation moins que parfaite observée. Ces modifications reposent sur l'hypothèse selon laquelle la réponse à un niveau fort à une oreille peut inhiber la réponse d'un son moins fort reçu par l'autre oreille, hypothèse déjà utilisée et discutée dans le modèle binaural de détection du signal de Breebaart *et al.* (2001). En terme de sonie binaurale, ce concept d'inhibition contralatérale a été introduit par Gigerenzer et Strube (1983).

L'implémentation des inhibitions contralatérales suggérée par Moore et Glasberg (2007) revient à appliquer un facteur d'atténuation $INHIB_G$ à la sonie spécifique calculée pour l'oreille gauche (N'_G) et un facteur d'atténuation $INHIB_D$ à la sonie spécifique calculée pour l'oreille droite (N'_D). Les étapes de calcul sont les suivantes :

- Le diagramme de sonie spécifique à chaque oreille (N'_G et N'_D) est au préalable lissé par convolution avec la fonction gaussienne W définie dans l'équation 1.10. Cela permet de représenter l'étalement de l'inhibition aux fréquences voisines, en accord avec les résultats de Scharf (1969) (un fort niveau à une fréquence à l'oreille gauche peut avoir un effet d'inhibition sur une fréquence différente présentée à l'oreille droite).

$$W(g) = \exp(-(Bg)^2) \quad (1.10)$$

où g est la déviation normalisée par rapport à E (valeur sur l'échelle des ERB_N) et B est un paramètre décrivant l'étalement de l'inhibition ($B = 0.08$).

- Les facteurs d'atténuation $INHIB_G$ et $INHIB_D$ sont déterminés pour chaque ERB à partir des diagrammes de sonie spécifique lissés $N'_G(E)_{liss}$ et $N'_D(E)_{liss}$:

$$INHIB_G(E) = 2 / [1 + (sech(N'_D(E)_{liss}/N'_G(E)_{liss}))^{1.5978}] ; \quad (1.11)$$

$$INHIB_D(E) = 2 / [1 + (sech(N'_G(E)_{liss}/N'_D(E)_{liss}))^{1.5978}] ; \quad (1.12)$$

L'évolution du facteur d'atténuation en fonction du rapport des sonies spécifiques $N'_{D_{liss}}/N'_{G_{liss}}$ est représentée dans la figure 1.11 pour chaque oreille.

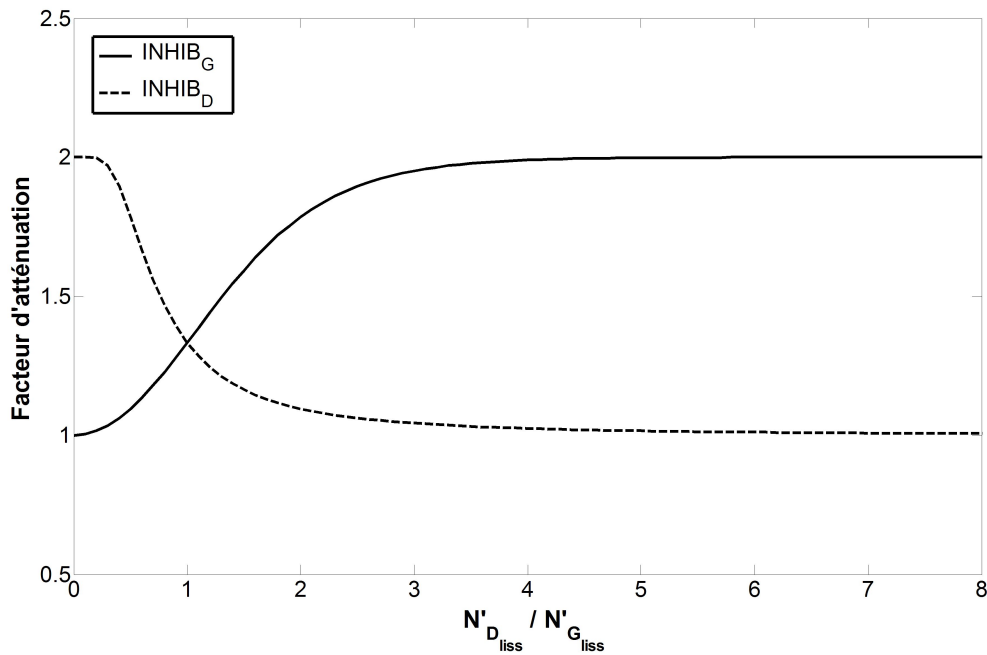


FIGURE 1.11: Facteur d'atténuation pour chaque oreille en fonction du rapport de sonies spécifiques $N'_{D_{liss}}/N'_{G_{liss}}$.

Ainsi, lorsque la sonie spécifique à gauche $N'_G(E)_{liss}$ est très grande devant la sonie spécifique à droite $N'_D(E)_{liss}$, le rapport $N'_D(E)_{liss}/N'_G(E)_{liss}$ tend vers zéro et le facteur d'atténuation appliqué à l'oreille gauche $INHIB_G$ tend vers 1 (pas d'inhibition contralatérale). Si le rapport $N'_D(E)_{liss}/N'_G(E)_{liss}$ augmente, l'atténuation appliquée à l'oreille gauche augmente également puis tend rapidement vers 2 (facteur d'atténuation déjà très proche de 2 lorsque le rapport des sonies spécifiques est égal à 4). Le raisonnement inverse est applicable pour l'atténuation appliquée à l'oreille droite (courbe en pointillé).

- L'atténuation correspondant aux inhibitions contralatérales est ensuite appliquée aux sonies spécifiques $N'_G(E)$ et $N'_D(E)$ par division (éq. 1.13 et 1.14).

$$N'_G(E)_{inhib} = N'_G(E) / INHIB_G(E) \quad (1.13)$$

$$N'_D(E)_{inhib} = N'_D(E) / INHIB_D(E) \quad (1.14)$$

- Finalement, comme dans le modèle classique la sonie est calculée pour chaque oreille en sommant les valeurs de sonie spécifique inhibée sur l'ensemble des ERBs. Puis la sonie globale binaurale est obtenue en sommant les sonies globales inhibées des deux oreilles :

$$N = N_{G_{inhib}} + N_{D_{inhib}} \quad (1.15)$$

Lorsque la présentation des sons est diotique, $N'_G(E)_{liss} = N'_D(E)_{liss}$ et donc $INHIB_G(E) = INHIB_D(E) = 4/3$ (intersection des courbes sur la figure 1.11). Cela correspond finalement à un rapport de sonie binaural-à-monastral de 1.5 ($N = 1.5N_G = 1.5N_D$), permettant selon Moore et Glasberg (2007) d'approcher le plus grand nombre de données expérimentales.

c) Modèle de Sivonen et Ellermeier (2008) :

Le modèle de sonie binaurale proposé par Sivonen et Ellermeier (2008) a été développé dans l'optique de prédire au mieux les données psychoacoustiques recueillies au sein d'une série d'études visant à mesurer la sonie de sources artificielles spatialisées dans un environnement réel (Sivonen *et al.* (2005), Sivonen et Ellermeier (2006), Sivonen (2007), Sivonen et Ellermeier (2008)). Comme dans le modèle de Moore et Glasberg (2007), la sommation de sonie binaurale est considérée moins que parfaite. En revanche, contrairement au modèle de Moore et Glasberg (2007) où la sommation de sonie binaurale était calculée à partir de la sonie estimée à chaque oreille, Sivonen et Ellermeier (2008) ont suggéré de modéliser la sommation binaurale directement à partir des spectres physiques mesurés à chaque oreille puis de calculer la sonie (modèle diotique). Les différentes étapes du modèle sont les suivantes :

- Les signaux de pression aux deux oreilles sont mesurés à l'aide d'une tête artificielle ($p_{2G}(t)$ et $p_{2D}(t)$).
- Les spectres d'amplitude des signaux aux oreilles sont analysés par bandes de tiers d'octave ($|P_{2G}(B_N)|$ et $|P_{2D}(B_N)|$ où B_N désigne la bande de tiers d'octave concernée).
- La sommation de sonie binaurale est ensuite effectuée dans chaque bande de tiers d'octave. Elle est modélisée par une règle de sommation logarithmique des niveaux en décibels gauche ($L_G(B_N)$) et droit ($L_D(B_N)$) définie dans l'équation 1.16.

$$L_{diot} = g \cdot \log_2 \left[2^{\frac{L_G(B_N)}{g}} + 2^{\frac{L_D(B_N)}{g}} \right] - g \quad (1.16)$$

où L_{diot} est le niveau diotique équivalent (correspondant au niveau à appliquer identiquement à chaque oreille pour avoir la même sonie) et g désigne le gain de sommation binaurale. La valeur $g = 3 \text{ dB}$ a finalement été retenue, permettant de prédire au mieux les données subjectives.

Cette règle de sommation binaurale des niveaux correspond à une sommation de puissance des signaux. En effet, si $|P_{2diot}(B_N)|$ désigne la pression diotique équivalente correspondant à L_{diot} , l'expression 1.16 devient alors :

$$|P_{2diot}(B_N)| = \frac{1}{\sqrt{2}} \sqrt{|P_{2G}(B_N)|^2 + |P_{2D}(B_N)|^2} \quad (1.17)$$

- Le spectre du signal diotique équivalent ainsi obtenu est ensuite modifié par l'application d'une HRTF inverse (correspondant à l'incidence frontale et mesurée sur la même tête artificielle que celle utilisée pour la mesure), afin de se ramener à un signal mesuré au centre de la tête en son absence ($p_1(t)$ de spectre $|P_1(B_N)|$).
- Finalement, la sonie du signal théorique $p_1(t)$ est calculée par un modèle de sonie diotique standard (e.g. ISO-532B (1975)).

En situation diotique d'écoute les niveaux L_G et L_D sont égaux. Le niveau monaural équivalent $L_{mon} = L_{diot} + g$ est donc supérieur de 3 dB aux niveaux à gauche et droite. Pour un son pur dont la fonction de sonie a un exposant $\alpha = 0.6$, cela correspond à un rapport de sonie binaural-à-monastral d'environ 1.2. D'autre part, comme pour les inhibitions contralatérales, plus la différence interaurale de niveau est grande, plus le poids attribué au niveau le plus fort (oreille ipsilatérale) est important dans le calcul de sonie par rapport au poids attribué à l'oreille contralatérale.

Remarque : Contrairement au modèle de Moore et Glasberg (2007) où la sommation binaurale se fait en sonie, dans le modèle de Sivonen et Ellermeier (2008), la sommation binaurale est calculée à partir des niveaux d'intensité, avant de passer dans le domaine perceptif. Or, d'un point de vue physiologique, la détermination de l'allure de l'excitation en fonction de la répartition de l'énergie spectrale (influence de la forme des filtres auditifs) ainsi que les phénomènes de masquage fréquentiel ont lieu au niveau périphérique (indépendamment pour chaque oreille). Ainsi, dans le cas où les signaux présentent des différences spectrales importantes entre les deux oreilles, il est possible que la sommation binaurale avant le calcul du diagramme d'excitation et de la sonie spécifique soit susceptible d'augmenter l'erreur de prédiction de la sonie globale. De ce point de vue, l'approche du modèle de Moore et Glasberg (2007) est plus complexe (car basée sur un plus grand nombre d'hypothèses) mais plus intuitive que celle du modèle de Sivonen et Ellermeier (2008).

1.4.2 Validité et limitations

1.4.2.1 Domaines de validité

Les trois modèles de sonie binaurale présentés sont relativement récents et leurs domaines de validité dans les situations dichotiques ou en présence de sources réelles ont été peu définis pour le moment. Quelques exemples d'études ayant confronté les modèles à des données subjectives sont présentés ici.

Schlittenlacher *et al.* (2011) ont montré que le modèle standard ANSI-S3.4 (2007) tendait à surestimer la sonie pour des bruits roses (égalisés en champ libre) dans le cas d'une présentation diotique des sons, ce qui n'était pas le cas du modèle DIN-45631 (1990) (version modifiée de l'ISO-532B (1975)). Ce résultat est en accord avec la comparaison des deux modèles déjà effectuée dans Fastl *et al.* (2009). En revanche, l'ANSI-S3.4 (2007) fonctionne mieux pour les niveaux faibles, probablement grâce à l'attention particulière portée au calcul de la sonie autour du seuil d'audition. Schlittenlacher *et al.* (2011) ont montré que l'ANSI-S3.4 (2007) surestimait également la sonie pour des sources de l'environnement considérées comme stationnaires (ventilateur d'ordinateur, bruit am-

biant urbain la nuit, aspirateur, sèche-cheveux, etc...). Dans les travaux de Boullet (2005), le modèle de Moore *et al.* (1997) (à l'origine de l'ANSI-S3.4 (2007)) avait déjà été testé sur des sons stationnaires environnementaux restitués en incidence frontale en champ libre (papier froissé, chalumeau, perceuse, flûte, vélo en roue libre, etc...). Les sons avaient été égalisés en sonie avec un son pur à 1 kHz lors d'une étude préalable (Meunier *et al.* (2000)). Or, le modèle de Moore *et al.* (1997) surestimait aussi systématiquement la sonie pour les sons au-dessus de 60 phones, contrairement au modèle de Zwicker. La surestimation était moins importante pour les sons dont le spectre se situait en basses fréquences. D'une manière générale, il a été mis récemment en évidence que les modèles DIN-45631 (1990) et ANSI-S3.4 (2007) surestimaient tous les deux la sommation de sonie spectrale au-delà de 1 kHz, expliquant en partie les résultats précédents (Schlittenlacher *et al.*, 2015). Concernant l'aspect dichotique, le modèle soutient l'hypothèse d'une sommation parfaite des sonies calculées à chaque oreille (e.g. Marks (1978)), allant ainsi à l'encontre de la plupart des données subjectives récentes soutenant l'hypothèse d'une sommation de sonie moins que parfaite.

Selon nos connaissances, le modèle proposé par Moore et Glasberg (2007) a été moins souvent testé dans la littérature. En revanche, les prédictions du modèle ont été confrontées aux données subjectives de plusieurs études de référence par les auteurs eux-mêmes (Moore et Glasberg, 2007). Ainsi, le modèle coïncide précisément avec les résultats de Scharf (1969) et Algorn *et al.* (1989a) dans des situations dichotiques d'écoute composées de sons purs de fréquences différentes à chaque oreille ou avec les résultats de Keen (1972) dans des situations dichotiques de niveau pour un son pur à 1 kHz. Il prédit également bien les résultats de l'expérience 1 de Zwicker et Zwicker (1991) montrant que la sonie binaurale (en situation dichotique de niveau) est indépendante du niveau global, de la fréquence centrale pour des bandes de bruit tiers d'octave ou de la forme du bruit (rouge, rose, bleu) pour des bruits large-bande. En revanche, le modèle de Moore et Glasberg (2007) a tendance à sous-estimer la sonie lorsque les sources sont spatialisées dans un environnement réel (Sivonen et Ellermeier, 2006). L'écart de prédiction de la sensibilité directionnelle en sonie est maximal (1.7 dB) pour une sinusoïde à 5 kHz spatialisée à 90° d'azimut. La sous-estimation est donc tout de même moins importante que pour le modèle de Moore *et al.* (1997) pour lequel la sous-estimation avoisinait les 5 dB dans les mêmes conditions (cf. Sivonen et Ellermeier (2006)). Les prédictions du modèle étaient également en désaccord avec les résultats de l'expérience 2 de Zwicker et Zwicker (1991), donc dans le cas où des différences spectrales importantes opèrent entre les deux oreilles. En effet, l'expérience consistait à diviser un bruit large-bande en sous-bandes distribuées alternativement à chaque oreille. Lorsque le nombre de sous-bandes alternées augmentait, la sonie diminuait, contrairement aux prédictions du modèle de Moore et Glasberg (2007). D'une manière générale, le modèle semble légèrement moins performant lorsque les stimuli s'approchent d'une situation réelle d'écoute (sources spatialisées, différences spectrales importantes entre les deux oreilles pour des bruits large-bande).

Le modèle de Sivonen et Ellermeier (2008) est un modèle de sommation binaurale basé essentiellement sur les données des différentes expériences menées par les auteurs pour une source artificielle (bande de bruit tiers d'octave ou large-bande) spatialisée en champ libre ou dans une salle d'écoute (Sivonen *et al.* (2005), Sivonen et Ellermeier (2006), Sivonen (2007), Sivonen et Ellermeier (2008)). Cependant, très peu d'études ont testé ce modèle à notre connaissance. Song *et al.* (2005) avaient mis en évidence une surestimation du modèle pour une source bitonale en incidence frontale en champ libre ou pour deux sources tonales spatialisées (surestimation également observée pour le modèle

de sonie diotique ISO-532B (1975)). Ce résultat met peut-être en avant une limite des modèles en présence de plusieurs sources tonales spatialisées dans un environnement réel.

Les prédictions des modèles de Moore *et al.* (1997), Moore et Glasberg (2007) et Sivonen et Ellermeier (2008) ont également été comparées pour des sources de bruit spatialisées dans un champ acoustique simulé plus ou moins diffus (Miranda et Cabrera, 2008). Le degré de diffusion du champ acoustique était caractérisé par la quantité de corrélation interaurale (cf. § 1.3.3 pour plus de détails). D'une manière générale, le modèle de Sivonen et Ellermeier (2008) (combiné avec le modèle diotique de Moore *et al.* (1997) plutôt que l'ISO-532B (1975)), fournissait une meilleure estimation de la sonie que les deux autres modèles. De plus, pour tous les modèles, lorsque la corrélation interaurale diminuait, la sonie en hautes fréquences (bande HF) avait tendance à être surestimée et la sonie du bruit rose (large-bande) avait tendance à être sous-estimée. Cela montre donc que la corrélation interaurale peut avoir une influence sur la sonie, pouvant altérer de manière non négligeable la prédiction des différents modèles si elle n'est pas prise en compte. On notera que dans Miranda et Cabrera (2008), le modèle de Moore *et al.* (1997) utilisé en configuration diotique à partir d'un enregistrement monaural au centre de la tête en son absence prédisait étonnamment mieux les données subjectives que les trois modèles binauraux présentés ci-dessus (ces derniers prenant pourtant en compte l'effet des différences interaurales de niveau). Ce résultat reste cependant difficile à interpréter.

Finalement, le modèle de Sivonen et Ellermeier (2008) est peut-être plus performant en présence de sources spatialisées, ayant été conçu dans ce contexte. Le modèle de Moore et Glasberg (2007) présente en revanche une approche plus représentative de la physiologie, prenant en compte de nombreux détails à chaque étape de calcul. Il a été testé dans des situations d'écoute moins réalistes mais plus variées. Le modèle de Moore *et al.* (1997) (ou ANSI-S3.4 (2007)) quant à lui ne tient pas compte des résultats récents rapportant une sommation de sonie binaurale moins que parfaite et a tendance à surestimer la sonie pour des sons large-bande et environnementaux. Il ne tient pas non plus compte de l'influence de l'ITD sur la sommation de sonie binaurale et a tendance au contraire à sous-estimer la sonie en présence de grands ILDs (Sivonen et Ellermeier, 2008). D'une manière générale, les modèles de sonie binaurale présentés ici sont basés sur des diagrammes d'excitation et de sonie spécifique calculés à partir des spectres d'amplitude des signaux mesurés aux oreilles. Il ne prennent donc pas en compte l'effet de la phase (Mulligan *et al.*, 1985), de la corrélation interaurale (Edmonds et Culling (2009), Miranda et Cabrera (2008)) ou des processus de plus haut niveau comme la constance perceptuelle en sonie (Zahorik et Wightman, 2001), le groupement de l'information (McAdams *et al.*, 1998) ou les interactions multisensorielles (Fastl, 2004). Il reste donc de nombreux aspects importants à comprendre et à intégrer dans les modèles de sonie binaurale afin d'améliorer la précision des estimations.

1.4.2.2 Influence du choix des HRTFs

Les modèles de sonie binaurale permettent de calculer la sonie d'un champ acoustique à partir des niveaux à l'oreille gauche et à l'oreille droite mesurés à l'aide d'une tête artificielle. Cependant, pour un champ acoustique donné, les niveaux mesurés aux oreilles peuvent varier selon le modèle de tête artificielle utilisé (variation suivant la forme de la tête, les matériaux et surtout la place des

microphones par rapport à l'entrée du conduit auditif). Le comportement acoustique de chaque tête artificielle peut être caractérisé par ses HRTFs.

Dans le modèle ANSI-S3.4 (2007), la fonction de transfert entre le champ acoustique (pression au centre de la tête en son absence) et le niveau au tympan correspond à une fonction de transfert moyenne représentative de l'humain adulte. Deux fonctions de transfert sont disponibles dans le modèle, correspondant à des mesures soit en champ libre soit en champ diffus. Ainsi une source frontale à 1 kHz produisant un niveau de pression en champ libre de $P_1 = 40 \text{ dB SPL}$ au centre de la tête en son absence, induit (d'après la fonction de transfert en champ libre) des niveaux $P_2 = 42.6 \text{ dB SPL}$ aux tympans. Le modèle est alors calibré pour que ces niveaux aux tympans produisent une sonie de 1 sone. En revanche, si cette même source frontale (1 kHz, 40 dB SPL) est mesurée à l'aide d'une tête artificielle ayant ses propres caractéristiques, les niveaux P_2 mesurés par les microphones seront probablement différents de 42.6 dB SPL et le modèle prédira donc une sonie différente de 1 sone. C'est pourquoi il est important de s'interroger sur l'influence que peuvent avoir les HRTFs de la tête artificielle utilisée pour la mesure sur la prédiction finale du modèle de sonie.

Selon cette idée, nous proposons d'évaluer l'erreur de prédiction induite (pour le modèle ANSI-S3.4 (2007)) lorsque la mesure binaurale du champ acoustique est réalisée à l'aide de la tête artificielle Cortex MK1 de notre laboratoire. La figure 1.12 compare les HRTFs considérées dans la norme ANSI-S3.4 (2007) avec celles de la tête artificielle Cortex MK1, en champ libre (incidence frontale) ou en champ diffus. Pour l'ANSI-S3.4 (2007), le gain représenté pour chaque tiers d'octave correspond en réalité au gain appliqué pour la fréquence centrale du tiers d'octave. En tenant compte de cette approximation, la différence entre ces deux fonctions de transfert (représentée par les barres grises) est à l'origine de l'erreur sur la sonie, dont l'ampleur dépend de la répartition de l'énergie spectrale du stimulus.

A titre d'exemple, l'erreur en sonie a été quantifiée pour les différentes sources d'incidence frontale suivantes, enregistrées avec la tête artificielle :

- **1 kHz** : bruit tiers d'octave centré à 1 kHz, mesuré en champ libre et en champ diffus, à 60 dB SPL au centre de la tête en son absence.
- **5 kHz** : bruit tiers d'octave centré à 5 kHz, mesuré en champ libre et en champ diffus, à 60 dB SPL au centre de la tête en son absence.
- **Bruit rose** : bruit rose, mesuré en champ libre et en champ diffus, à 60 dB SPL au centre de la tête en son absence.
- **V1 et V2** : sons de ventilation mesurés à l'intérieur d'une voiture, à un niveau moyen de 53 dB(A) "aux tympans". Stimuli V1 V2 du chapitre 5, le spectre est fourni dans l'annexe E.
- **S1 et S2** : sources aéroacoustiques générées par une voiture en soufflerie (enregistrement à l'intérieur), à un niveau moyen de 53 dB(A) "aux tympans". Stimuli S1 et S2 du chapitre 5, le spectre est fourni dans l'annexe E.
- **A1** : son d'aspirateur, restitué sur une enceinte, mesuré dans une salle de cours du laboratoire,

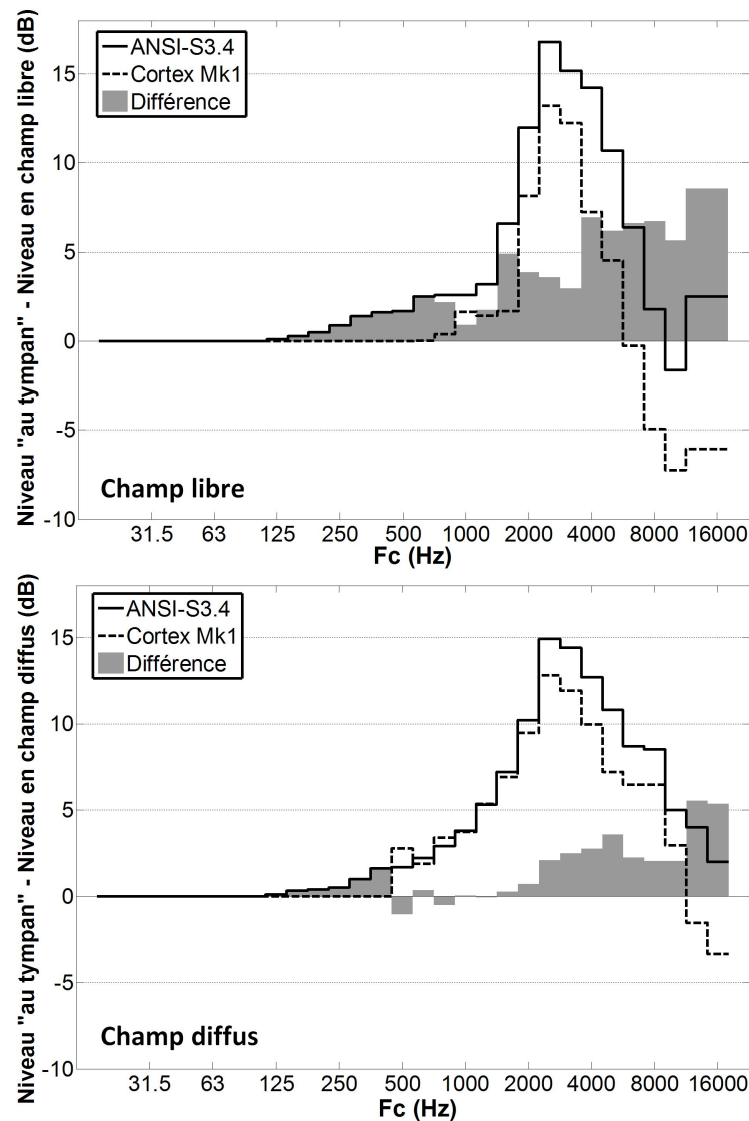


FIGURE 1.12: Niveaux par bandes de tiers d'octave correspondant aux HRTFs considérées dans la norme *ANSI-S3.4* (trait continu) et mesurées sur la tête Cortex Mk1 (trait en pointillé), en champ libre en incidence frontale (en haut) et en champ diffus (en bas). L'axe des ordonnées correspond donc à la différence entre le niveau "au tympan" et le niveau au centre de la tête en son absence. Les barres grises représentent la différence de niveau entre les deux fonctions de transfert (ANSI-S3.4 et Cortex MK1).

à un niveau moyen de 53 dB(A) "aux tympans". Stimulus *A1* du chapitre 5 (le spectre est fourni dans l'annexe E).

- **M1** : son de machine à laver, restitué sur une enceinte, mesuré dans une salle de cours du laboratoire, à un niveau moyen de 53 dB(A) "aux tympans". Stimulus *M1* du chapitre 5 (le

spectre est fourni dans l'annexe E).

La quantification de l'erreur consistait à comparer la sonie (N) calculée par le modèle ANSI-S3.4 (2007) directement à partir de l'enregistrement binaural, et la sonie (N_{Corr}) calculée en appliquant une correction au préalable, correspondant à la différence des fonctions de transfert (barres grises sur la figure 1.12). La figure 1.13 représente le rapport N_{Corr}/N pour les différentes sources testées. Pour les trois premières sources (1 kHz, 5 kHz et bruit rose) mesurées en champ libre et en champ diffus, l'erreur de sonie a été estimée en considérant respectivement les HRTFs en champ libre (cercles), et les HRTFs en champ diffus (carrés). Il en est de même pour les sources V1, V2, S1 et S2, étant donnée l'ambiguïté sur la nature du champ dans un habitacle de voiture en présence d'une multitude de source. En revanche, pour les sources A1 et M1 mesurées dans une salle de cours du laboratoire ($TR_{60} = 1.14$, $V = 162.4m^3$), le champ pouvait être considéré comme diffus au-delà de 167 Hz (selon la formule de Schroeder). Or, comme l'effet des HRTFs est quasiment inexistant en-dessous de cette fréquence, seule la correction correspondant au champ diffus a été considérée.

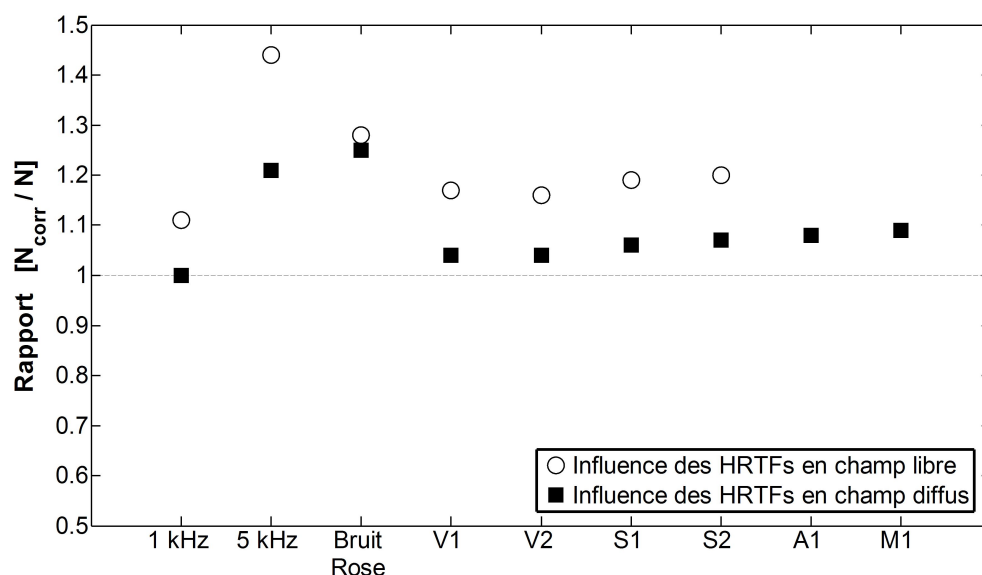


FIGURE 1.13: Rapport N_{Corr}/N représentant l'erreur de sonie induite par la différence entre les HRTFs de la tête artificielle Cortex Mk1 utilisée pour la mesure et les HRTFs prises en compte dans le modèle ANSI-S3.4 (2007) utilisé pour le calcul de sonie. L'influence des HRTFs sur la sonie est quantifiée pour différentes sources, en champ libre (cercles blancs) ou en champ diffus (carrés noirs).

Plus le rapport de sonie N_{Corr}/N est proche de 1 et plus l'erreur de prédiction de la sonie est faible. Une valeur du rapport supérieure à 1 représente une sous-estimation de la sonie si aucune correction n'est prise en compte. On constate d'une manière générale que l'utilisation de la tête artificielle Cortex Mk1 pour l'enregistrement entraîne une erreur de prédiction du modèle plus importante en champ libre qu'en champ diffus. Cela s'explique simplement par le fait que les différences physiques entre HRTFs sont généralement plus marquées en champ libre (cf. figure 1.12).

Pour les sources artificielles à 1 kHz et 5 kHz, l'erreur (dépendant uniquement des différences

d'HRTFs dans le tiers d'octave concerné) peuvent être importantes : jusqu'à 44% pour la source à 5 kHz en champ libre. Pour un bruit rose (même énergie par bande de tiers d'octave sur l'ensemble du spectre), la sonie est sous-estimée d'environ 25%, quelle que soit la nature du champ acoustique. Cela laisse donc présager une erreur encore plus importante si l'énergie spectrale de la source est concentrée d'avantage en hautes fréquences. Enfin, pour les sources réelles (V1, V2, S1, S2, A1 et M1), l'erreur est globalement plus faible que pour le bruit rose (particulièrement en champ diffus). Cela s'explique par la concentration plus importante de l'énergie dans les basses fréquences, là où les différences d'HRTFs interviennent peu. En champ libre, la sous-estimation de la sonie reste tout de même comprise entre 15% et 20%.

Cette analyse montre donc l'importance de prendre en compte les HRTFs de la tête artificielle utilisée pour la mesure, considération d'autant plus justifiée que le champ acoustique s'éloigne d'un champ diffus. Au regard de la précision des modèles de sonie, une erreur de 25% (pour le bruit rose) est déjà importante, elle correspond à une différence de niveau d'isophonie de 2.8 phones. Pour tout enregistrement binaural réalisé à l'aide d'une tête artificielle (ou aux oreilles d'un auditeur réel), il est donc vivement recommandé d'appliquer une correction avant de procéder au calcul de sonie. Cette correction correspond à la différence entre les HRTFs (ou autre correction de champ) définies dans le modèle de sonie utilisé (ANSI-S3.4 (2007) dans notre exemple) et celles correspondant au dispositif de mesure (pour le champ libre ou pour le champ diffus).

De la même manière, dans l'avant-dernière étape du modèle de Sivonen et Ellermeier (2008) consistant à se ramener au niveau de pression au centre de la tête en son absence (cf. figure 1.9), il est important d'utiliser les HRTFs (mesurées en champ libre ou en champ diffus selon les conditions expérimentales) de la tête artificielle ayant servi à faire la mesure. Dans la dernière étape consistant à calculer la sonie à l'aide du modèle ISO-532B (1975) (ou ANSI-S3.4 (2007)), il faut également veiller à utiliser l'égalisation proposée dans le modèle (champ direct ou champ diffus), en adéquation avec l'étape précédente et la mesure initiale.

On notera cependant que dans l'ensemble de la thèse, la performance des modèles a été évaluée en calculant des rapports de sonie entre deux sons égalisés au PSE, de contenus spectraux analogues. L'erreur réalisée sur la valeur de la sonie est donc similaire pour les deux sons et par conséquent, influe peu sur le rapport de sonie final (des sons au PSE). En revanche, si l'intérêt est porté directement sur la valeur de sonie estimée en sone, il est important de prendre en compte la correction. C'est donc ce qui a été envisagé dans la partie 4.6.2 du chapitre 4 concernant la prédiction de la sonie absolue d'une source bitonale.

Première partie

Sonie directionnelle et variabilité interindividuelle

Mesures de sonie directionnelle pour une source de bruit à bande étroite

Ce chapitre est principalement consacré à la réplication sur un plus grand nombre d'auditeurs (17) de l'expérience menée par Sivonen et Ellermeier (2008), dans le but principal de mieux quantifier et comprendre l'origine des différences interindividuelles rapportées par ces auteurs. Cette expérience consistait à comparer la sonie d'une source artificielle spatialisée dans un environnement réel à la sonie d'une source frontale de référence. Les mesures ont été réalisées au casque, en simulant les sources suivant différentes directions à l'aide du processus de synthèse binaurale générique. Le choix des stimuli ainsi que la méthode ont été directement tirés des différents travaux menés par Sivonen et Ellermeier (Sivonen et Ellermeier, 2006; Sivonen, 2007; Sivonen et Ellermeier, 2008).

2.1 Méthode

2.1.1 Auditeurs

Dix-sept auditeurs dont neuf hommes et huit femmes, âgés de 24 à 49 ans (moyenne d'âge : 27 ans), ont été sollicités parmi les étudiants de l'INSA de Lyon et le personnel du laboratoire pour participer à cette expérience. Sept des auditeurs (MV, EP, TL, CM, LB, LC et ML) pouvaient être considérés comme normo-entendants (seuils < 20 dB HL). Les dix autres auditeurs présentaient des pertes auditives légères unilatérales ou bilatérales (seuils ≤ 25 dB HL) à 1 kHz, 4 kHz ou 6 kHz (fréquences proches de celles des stimuli utilisés dans l'étude). En particulier, l'auditeur RS avait un seuil de 30 dB HL à l'oreille gauche à 1 kHz, l'auditeur SB un seuil de 40 dB HL à l'oreille droite à 6 kHz et l'auditeur WA un seuil de 30 dB HL à l'oreille droite à 4 kHz. Le détail des audiogrammes est présenté dans l'annexe A. Les auditeurs étaient indemnisés à hauteur de 10 euros par heure de test réalisée.

2.1.2 Stimuli

Les stimuli utilisés étaient des bruits roses monophoniques filtrés par bandes de tiers d'octave centrées à 1 kHz et 5 kHz (en accord avec les spécifications du standard ANSI-S1.11 (1986)). La durée des bruits était de 1 s, incluant une croissance et une décroissance en cosinus de 20 ms au début et à la fin du son (utilisation d'une fenêtre de Hann).

Les stimuli étaient ensuite convolués par des filtres (gauche et droit) permettant de réaliser la synthèse au casque de sources virtuelles (à environ 2 m du centre de la tête de l'auditeur) pour différentes directions de l'espace. Le principe général de la synthèse binaurale était de restituer aux oreilles de l'auditeur, les signaux qui y seraient parvenus en présence d'une source réelle en champ

libre. Cela a nécessité entre autre la mesure de réponses impulsionnelles binaurales (RIB) sur la tête artificielle Cortex Mk1 (TA) et des fonctions de transfert du casque positionné sur la TA. Le processus complet de la synthèse des stimuli est détaillé dans l'annexe B.

La méthode faisait appel à deux types de stimuli : un son frontal de référence et un son de comparaison spatialisé dans une direction donnée de l'espace (cf. § 2.1.3). La direction du son frontal de référence correspondait au couple de coordonnées azimuth/élévation ($0^\circ, 0^\circ$). Les différentes directions explorées pour le son de comparaison étaient constituées par l'ensemble des combinaisons entre les azimuths $0^\circ, 30^\circ, 60^\circ, 90^\circ, 150^\circ, 180^\circ$ (main gauche) et les élévations 0° et 45° . La position ($0^\circ, 90^\circ$) a également été ajoutée et correspondait à une source située au zénith de l'auditeur. On notera cependant que la position ($0^\circ, 0^\circ$) pour le son de comparaison a été testée pour la fréquence centrale 1 kHz afin d'avoir une mesure de référence, mais que cette direction n'a pas fait l'objet d'une nouvelle mesure à 5 kHz. Un certain nombre de directions ont ainsi été ajoutées par rapport à celles utilisées dans Sivonen (2007) et Sivonen et Ellermeier (2008), notamment toutes les positions d'élévation non nulle en dehors du plan sagittal.

L'intervalle inter-stimuli (ISI) était fixé à 500 ms entre son de référence et son de comparaison.

Le niveau global du son frontal de référence était fixé à 65 dB SPL pour une source mesurée en champ diffus. Un niveau global plus faible (45 dB SPL) avait également été étudié dans Sivonen et Ellermeier (2006), mais l'effet de la direction sur la sonie étant plus marqué à 65 dB SPL, seul ce niveau a été conservé. La méthode de calibration du casque est détaillée dans l'annexe C.

2.1.3 Procédure

Une procédure adaptative à choix forcé sur deux intervalles, suivant une règle 1-down/1-up a été mise en place afin d'égaliser la sonie de chaque source d'incidence non-nulle avec la sonie de la source frontale de référence. Une telle procédure permettait d'estimer le point à 50% de "réponses positives" sur la fonction psychométrique (Levitt (1971)), correspondant au point d'égalité subjective (PSE).

Chaque essai consistait à présenter successivement un son de comparaison d'incidence non-nulle (de niveau variable) et un son frontal de référence (de niveau fixe). L'ordre de présentation des deux sons était tiré de manière aléatoire à chaque essai afin d'éviter les erreurs systématiques de type "Time-order errors" (Hellström, 1979). Il était alors demandé à l'auditeur de désigner le son qu'il percevait comme étant le plus fort (1 ou 2). La consigne était formulée de la manière suivante : *Écoutez attentivement les paires de sons proposées. Pour chaque paire de sons, nous vous demandons de désigner le son (1 ou 2), que vous percevez comme étant le plus fort. L'évaluation doit être portée sur le son dans son ensemble.* Selon la réponse de l'auditeur, le niveau du son de comparaison était alors augmenté (si perçu plus fort) ou diminué (si perçu moins fort). Le pas de variation était fixé à 4 dB jusqu'à la 2^e inversion, à 2 dB jusqu'à la 4^e inversion puis à 1 dB jusqu'à la fin de la mesure. On notera que le niveau relatif du son de comparaison par rapport au son de référence était limité dans l'intervalle [-15 dB, +15 dB] afin d'éviter toute atteinte à la santé des auditeurs.

Une mesure se terminait lorsque le nombre de dix inversions était atteint. Le niveau d'égalisation était finalement estimé par la moyenne arithmétique du niveau relatif du son de comparaison (par rapport au son de référence) sur les six dernières inversions.

En tout, il y avait 25 conditions définies par les deux fréquences centrales (1 kHz et 5 kHz) et 12 directions (à 1 kHz, le son de référence frontal était également comparé à lui même afin d'obtenir une mesure de référence). Pour chaque condition, deux mesures étaient effectuées avec des niveaux relatifs de départ différents pour le son de comparaison (+10 dB ou -10 dB). Chaque mesure était ensuite répétée quatre fois, menant à un total de 200 mesures par auditeur à effectuer. L'estimation du PSE pour chaque condition, était finalement définie comme la moyenne des 8 niveaux d'égalisation obtenus (2 niveaux de départ \times 4 répétitions).

Les mesures ont été réparties en 25 blocs de 8 mesures, de manière pseudo-aléatoire (en limitant à 2 le nombre de répétitions d'une même condition au sein d'un même bloc). Les 8 mesures de chaque bloc étaient entrelacées de manière aléatoire. Ainsi, pour chaque présentation, la paire de sons présentée correspondait à une des 8 mesures du bloc, sélectionnée au hasard.

L'ordre de présentation des 25 blocs a également été tiré au sort pour chacun des auditeurs. Le test était ensuite réparti en 7 séances de 1h (7h au total). La première séance était réservée à l'explication de la consigne et la réalisation des 3 premiers blocs (1h). Ensuite, 4 blocs étaient réalisés par séance pendant 5 séances (5h). La dernière séance consistait à réaliser les 2 blocs restants et à effectuer si besoin, des mesures supplémentaires (1h). Une heure de test supplémentaire a été nécessaire pour 3 des auditeurs.

Critères de validité Inter/Intra mesure Afin d'avoir un contrôle sur la pertinence des niveaux d'égalisation estimés, deux critères arbitraires ont été fixés de manière empirique :

- Une mesure était considérée comme étant valide si l'écart-type du niveau du son de comparaison sur les 6 dernières inversions était inférieur ou égal à 2.5 dB (2.5 fois le pas de variation final). Si ce n'était pas le cas, une nouvelle mesure était réalisée jusqu'à ce que le critère soit vérifié.
- D'autre part, les 4 répétitions de la même mesure étaient considérées comme étant consistantes entre elles si l'écart-type des 4 niveaux d'égalisation obtenus était inférieur ou égal à 1.5 dB (1.5 fois le pas de variation final). Si ce critère n'était pas vérifié, des répétitions de la mesure étaient ajoutées pour cette condition, jusqu'à ce que le critère de validité entre les mesures soit vérifié.

Au total cela a représenté 1 à 4 blocs supplémentaires pour 6 auditeurs et 9 blocs supplémentaires pour l'auditeur LC ayant visiblement eu plus de difficulté à effectuer la tâche.

2.1.4 Dispositif expérimental

La procédure adaptative a été implémentée sous Matlab. Les sons étaient générés à l'aide d'une carte son Echo Gina 24 (16 bits, 44100 Hz) puis restitués à l'aide d'un casque Sennheiser HD 650 dans une cabine audiométrique. Un écran était utilisé pour l'affichage de l'interface visuelle du test (rappel de la consigne, réponse,...). Les auditeurs fournissaient leur réponse via les touches "1" et "2" d'un clavier customisé pour l'expérience (aucune autre touche disponible).

2.1.5 Mesure des niveaux SPL

Dans le but de relier les mesures de sonie aux niveaux SPL "aux tympans" (G et D), et ainsi prendre en compte la contribution des HRTFs dans le jugement de sonie binaurale, les niveaux SPL ont été mesurés sur la tête artificielle pour tous les stimuli utilisés dans cette expérience.

Pour les deux fréquences centrales, et pour l'ensemble des angles d'incidence, les stimuli ont été joués au casque positionné sur la tête artificielle. Le casque était calibré à un niveau global correspondant à 65 dB SPL en champ diffus pour la source frontale de référence (cf. annexe C). Les niveaux SPL "aux tympans" correspondants ont ainsi été mesurés à l'aide des microphones de la TA. Comparés à 65 dB SPL, ces niveaux correspondent en théorie aux gains engendrés par les HRTFs de la TA pour les bandes de tiers d'octave concernées.

Pour chaque angle d'incidence, ces niveaux ont ensuite été normalisés par rapport aux niveaux SPLs "aux tympans" mesurés pour la référence frontale. Les résultats obtenus sont présentés dans la section 2.2.3 (Tab. 2.1).

Notons qu'ici nous avons utilisé des RIB génériques, donc les variations d'ILDs avec l'angle d'incidence étaient les mêmes pour tous les auditeurs. Cependant, la réponse du casque utilisé était également générique, ce qui implique que les niveaux SPL mesurés sur la TA étaient une approximation des niveaux réellement reçus par les différents auditeurs. Mais les stimuli étant des bruits à bande étroite, cette erreur concernait uniquement le niveau global absolu à chaque fréquence centrale (puisque la réponse du casque ne dépendait pas de l'angle d'incidence). Ainsi, les niveaux SPL relatifs (par rapport à la référence frontale) restaient donc en théorie invariants d'un auditeur à l'autre.

2.2 Résultats

Sensibilités directionnelles en sonie L'ensemble des résultats de cette expérience sont exprimés sous la forme de "sensibilités directionnelles en sonie" (DLS), notion introduite par Sivonen et Ellermeier (2006). La "sensibilité en sonie" correspond à l'opposé du niveau relatif du son de comparaison mesuré au point d'égalité subjective (PSE) en sonie et l'expression "sensibilités directionnelles en sonie" désigne l'évolution des sensibilités en sonie en fonction de l'angle d'incidence. Ainsi, une valeur de DLS positive dans une direction donnée correspondra à une augmentation de la sensibilité en sonie dans cette direction (le son de comparaison a été perçu plus fort que la référence frontale). Inversement, une DLS négative traduira une diminution de la sonie par rapport à l'incidence frontale.

2.2.1 Résultats individuels

Les DLS obtenues pour chacun des 17 auditeurs ayant participé à l'expérience sont représentées dans la figure 2.1 en fonction de l'angle d'incidence. La partie de gauche représente les résultats obtenus pour la fréquence centrale $F_c = 1 \text{ kHz}$, tandis que la partie de droite représente les résultats obtenus pour la fréquence centrale $F_c = 5 \text{ kHz}$. La dimension verticale correspond à la variation de l'angle d'élévation ($\Theta = 0^\circ$ en haut, $\Theta = 45^\circ$ au milieu, $\Theta = 90^\circ$ en bas).

Concernant la condition témoin à 1 kHz, qui consistait à égaliser en sonie la référence frontale avec elle-même, les DLS individuelles varient entre -0.52 et 1.14 dB. Un test des signes a permis

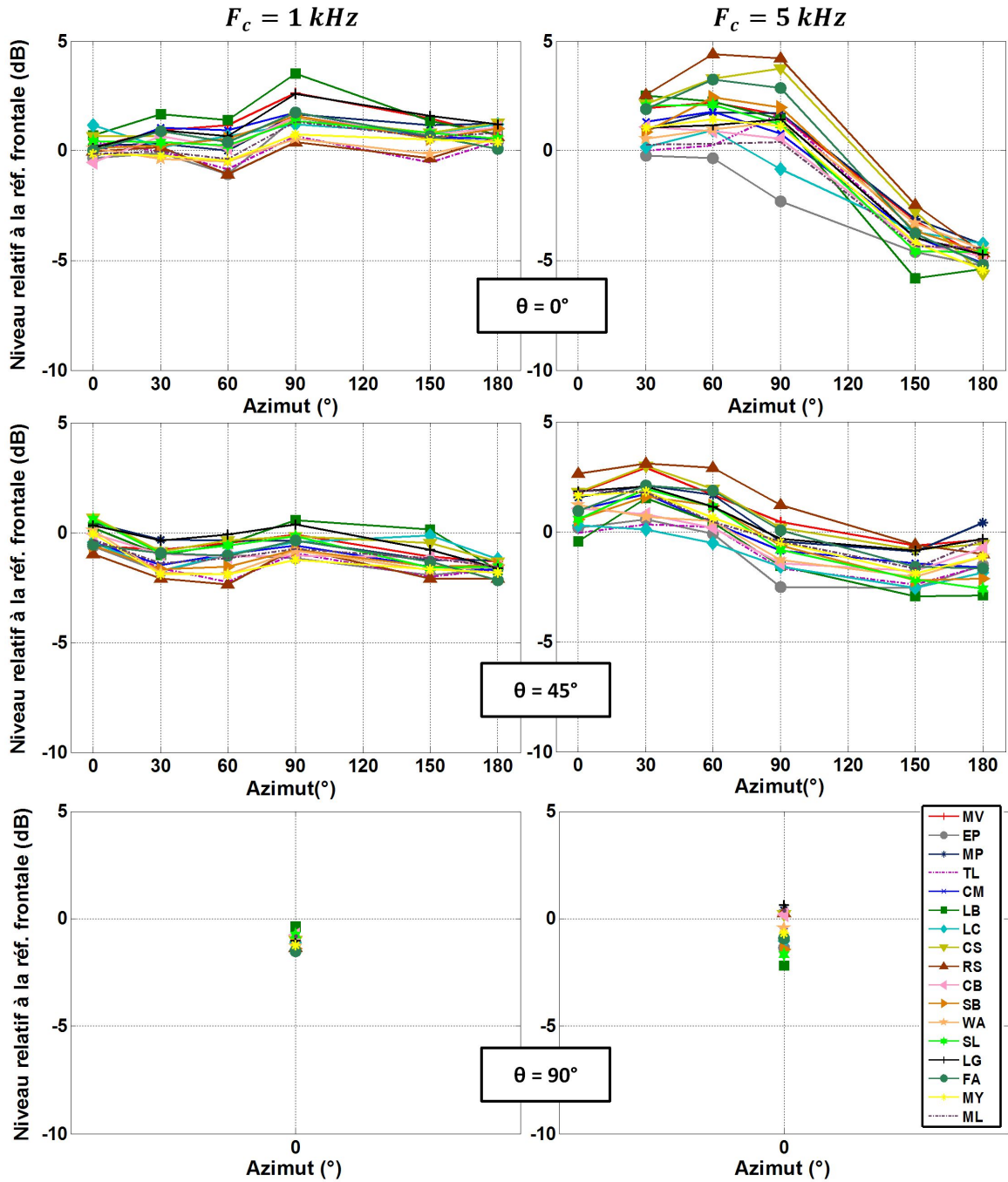


FIGURE 2.1: Sensibilités directionnelles en sonie (DLS) individuelles pour les 17 auditeurs en fonction de l'azimut. Résultats pour les deux fréquences centrales $F_c = 1 \text{ kHz}$ (à gauche) et $F_c = 5 \text{ kHz}$ (à droite) et pour les trois élévations $\Theta = 0^\circ$ (en haut), $\Theta = 45^\circ$ (au milieu) et $\Theta = 90^\circ$ (en bas).

de montrer que la médiane des DLS individuelles pour cette condition (égale à 0.1 dB) n'était pas différente de zéro au sens statistique ($p=0.33$). Cela confirme donc le bon fonctionnement de la méthode de mesure.

On notera également que conformément au critère de validation entre les mesures (cf. § 2.1.3), la variabilité des différentes répétitions d'une même mesure est faible (écarts-types toujours inférieurs ou égaux à 1.5 dB). Le jugement des auditeurs est donc cohérent, malgré la difficulté apparente de la tâche. De plus, une ANOVA à mesure répétée a révélé l'absence d'effet des différents facteurs (fréquence, azimuth, élévation) sur cette variabilité intra-auditeur. Cela montre donc qu'il n'y a pas de conditions en particulier pour lesquelles la tâche aurait été plus difficile pour les auditeurs.

D'une manière générale, on constate que la sonie varie en fonction de l'angle d'azimut dans le plan horizontal et à 45° d'élévation. L'évolution de la sonie avec l'azimut varie avec l'élévation et avec la fréquence. A 5 kHz les variations de sonie dans le plan horizontal sont plus importantes que dans le plan à 45°, mais aussi plus importantes que dans le plan horizontal à 1 kHz. Ce schéma général des DLS est constaté pour chaque auditeur indépendamment.

Cependant il existe des différences interindividuelles importantes. Sur l'ensemble des angles d'incidence, à 1 kHz, la plus grande variation de DLS est de 5.2 dB (auditeur LB) et la plus petite variation est de 2.5 dB (WA) ; à 5 kHz, la plus grande et la plus petite variation sont respectivement de 9.3 dB (CS) et 5.1 dB (LC). L'amplitude de variation de la DLS sur l'ensemble des angles d'incidence peut donc varier du simple au double d'un auditeur à l'autre. Cela se traduit par des différences interindividuelles parfois importantes en sonie pour un angle d'incidence donné. La dispersion est globalement plus importante à 5 kHz qu'à 1 kHz. A 5 kHz, ces différences sont les plus grandes dans le plan horizontal pour l'azimut 90° avec un écart de 6.5 dB entre les DLS des deux auditeurs extrêmes RS et EP. L'écart entre ces deux auditeurs est de 4.7 dB à 60° d'azimut. On peut remarquer que les deux auditeurs EP et LC se démarquent à 90° d'azimut en tirant les valeurs vers le bas. Ce sont d'ailleurs les seuls dont la DLS est négative pour cette condition. A titre de comparaison, une différence de 6 dB entre les deux auditeurs extrêmes avait été observée par Sivonen et Ellermeier (2008) à 90° d'azimut dans le plan horizontal pour un bruit centré à 5 kHz.

2.2.2 Groupes d'auditeurs

Comme l'ont montré les résultats précédents, il existe des différences parfois grandes entre les sensibilités en sonie des différents auditeurs pour un angle d'incidence donné et entre les amplitudes de variation de la sonie suivant les différents angles d'incidence. Afin de comprendre un peu mieux la façon dont s'organisent ces différences, il nous a semblé intéressant de rechercher les possibilités éventuelles de regroupements entre les auditeurs.

La figure 2.2 propose une classification hiérarchique ascendante des différents auditeurs par application de la méthode de Ward (classification basée sur les distances euclidiennes). Le calcul des distances a été effectué directement à partir des niveaux d'égénéralisation individuels, estimés pour chaque condition expérimentale.

La lecture de l'arbre suggère la formation de deux principaux groupes d'auditeurs : un groupe A contenant 7 auditeurs (TL, WA, CB, ML, MY, EP et LC) et un groupe B contenant 10 auditeurs

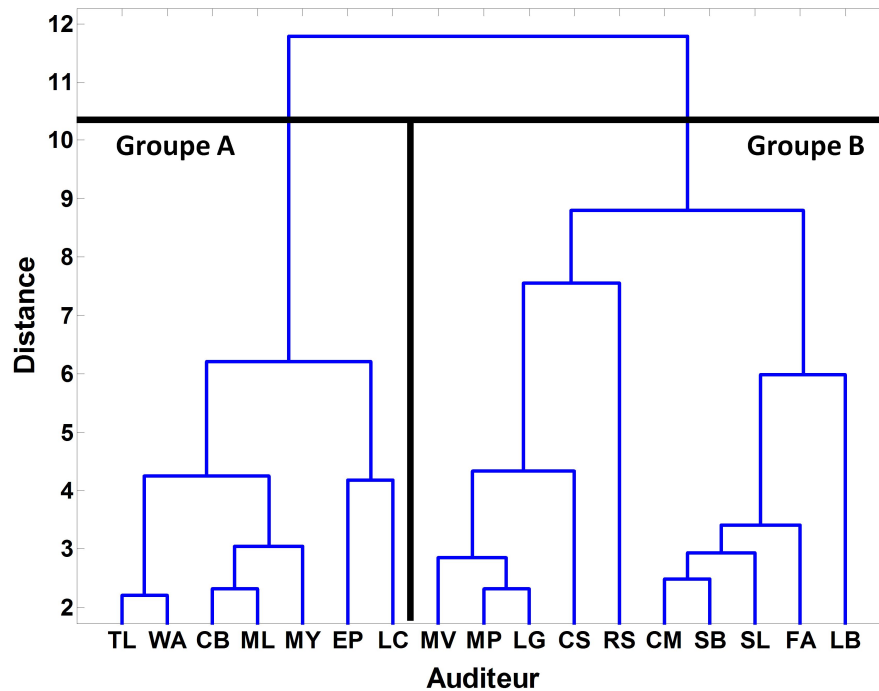


FIGURE 2.2: Dendrogramme de classification hiérarchique des différents auditeurs au regard de la DLS, basée sur les distances euclidiennes.

(MV, MP, LG, CS, RS, CM, SB, SL, FA et LB).

La figure 2.3 représente les DLS médianes ainsi que les écarts interquartiles associés pour chaque groupe d'auditeurs A et B. Elle est organisée de la même manière que la figure 2.1, à l'exception de l'élévation 90° qui n'y est pas représentée (apportant peu d'information supplémentaire). En effet lorsque la source est au zénith, il n'y a pas de différence entre les deux groupes, les données recueillies pour cette incidence ne jouent donc pas un rôle déterminant dans la classification.

D'après la figure, la différence entre les deux groupes est unilatérale. En effet, les auditeurs du groupe A ont des DLS systématiquement plus faibles que ceux du groupe B. Ces différences évoluent peu avec l'azimut ou l'élévation mais sont globalement plus marquées à 5 kHz qu'à 1 kHz. On notera également que dans la majorité des cas les différences entre les groupes ne sont plus discernables dans le plan sagittal (azimuts 0° et 180°).

Les différences entre les groupes ont été analysées à l'aide d'une ANOVA de Kruskal-Wallis. Les résultats du test ont montré qu'en dehors du plan sagittal, les différences entre les groupes A et B étaient significatives pour 13 conditions sur 16 ($p < 0.05$), la significativité de l'effet étant plus importante à 5 kHz ($p < 0.02$). Pour toutes les incidences d'azimut 0 ou 180° , il n'y a pas de différence entre les groupes.

Les dimensions à l'origine de la classification permettraient donc de discerner les groupes en dehors du plan sagittal et de façon plus marquée à 5 kHz, soit a priori lorsque les différences interaurales sont plus grandes. Il est donc vraisemblable que cette classification illustre finalement des stratégies

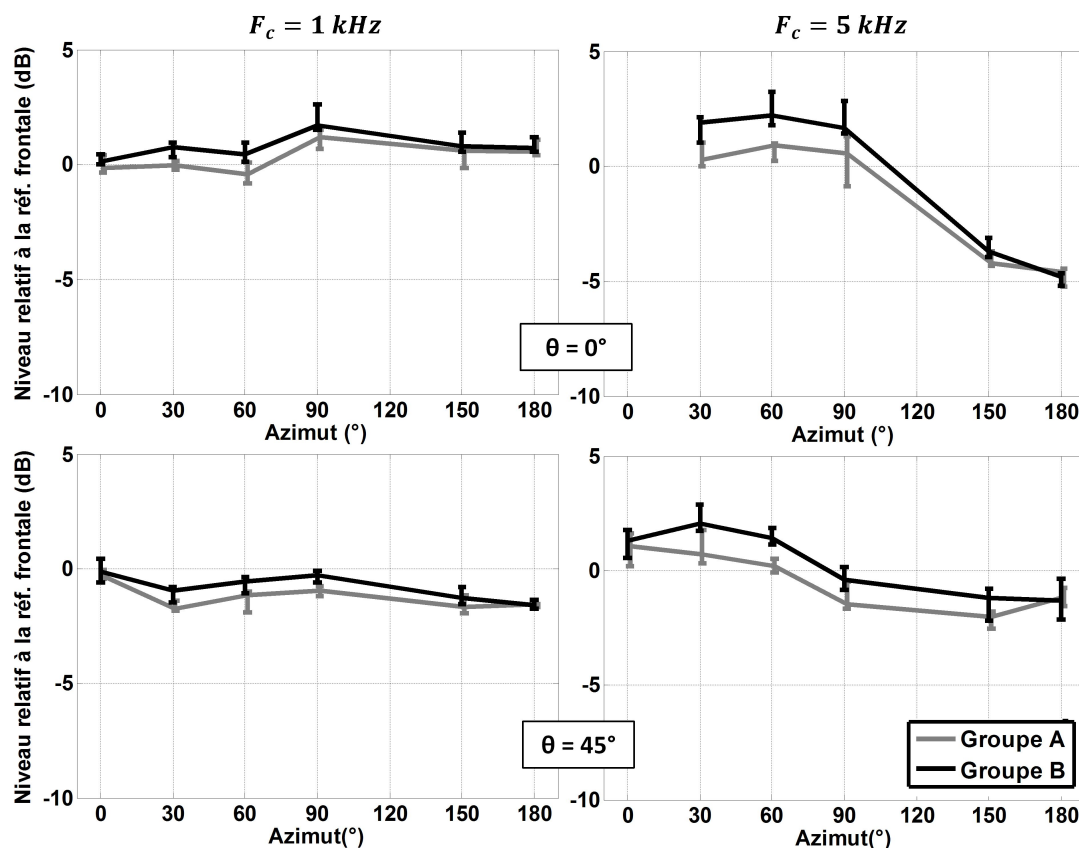


FIGURE 2.3: DLS médianes et écarts interquartiles en fonction de l'azimut pour chaque groupe d'auditeur (Groupe A en vert, Groupe B en noir).

individuelles différentes de sommation des niveaux SPL. Il est de plus particulièrement intéressant de noter la stabilité des groupes d'auditeurs sur l'ensemble des conditions expérimentales.

L'écart entre les médianes des deux groupes reste cependant faible au regard de la précision de la méthode d'une part et de la variation intra-individuelle d'autre part. En effet, l'écart entre les médianes est toujours inférieur à 1 dB à 1 kHz et inférieur à 1.5 dB à 5 kHz. Or, le pas de variation de la méthode est de 1 dB à la fin d'une mesure et il est fréquent que pour un auditeur donné, l'écart-type sur les différentes répétitions d'une même mesure soit compris entre 1 et 1.5 dB.

On retiendra donc la possibilité de discerner deux groupes d'auditeurs différents lorsque les différences interaurales sont grandes. Les groupes se distinguent par des niveaux de sensibilité en sonie systématiquement différents (niveau "fort" ou niveau "faible") et leur organisation est relativement stable sur l'ensemble des angles d'incidence. Cette classification reflète donc vraisemblablement l'existence de stratégies de sommation binaurale différentes. Cependant, les différences observées restent faibles, donc pour la suite de l'analyse, tous les résultats seront pris en compte de manière équivalente et la distinction entre les deux groupes d'auditeurs ne sera pas faite.

2.2.3 DLS médianes et niveaux SPL "aux tympans"

Niveaux SPL "aux tympans" :

Élévation	Azimut	1 kHz			5 kHz		
		G	D	ILD	G	D	ILD
0°	0°	0	0.1	-0.1	-	-	-
0°	30°	2.4	-3	5.4	3.3	-5.9	9.2
0°	60°	3.1	-4.7	7.8	4.5	-9.2	13.7
0°	90°	3.9	-2.4	6.3	4.3	-12.7	17
0°	150°	3.3	-3.6	6.9	-3.2	-6.4	3.2
0°	180°	1.1	1	0.1	-4.4	-4.5	0.1
45°	0°	0.5	0	0.5	-0.3	-0.9	0.6
45°	30°	0.9	-4.1	5	3.2	-5.2	8.4
45°	60°	1.6	-6.2	7.8	2.8	-9.1	11.9
45°	90°	1.7	-5.1	6.8	1	-11.2	12.2
45°	150°	0.1	-4	4.1	-1	-6.4	5.4
45°	180°	-1.5	-1.8	0.3	-3.2	-2.6	-0.6
90°	0°	-0.5	-1.3	0.8	-2.4	-2.5	0.1

Tableau 2.1: Niveaux SPLs relatifs mesurés sur la TA pour les différents angles d'incidence à 1 kHz et 5 kHz. Les ILDs associés ont également été normalisés par rapport aux ILDs des sources de référence frontales à chaque fréquence.

Il est admis qu'une importante partie de la variation de la sensibilité en sonie avec l'élévation et l'azimut peut-être expliquée par la variation des différences interaurales en fonction de l'angle d'incidence (induites par les HRTFs). Pour cette raison, afin de pouvoir relier les variations dans la DLS aux variations d'ILD, les niveaux de pression SPL "aux tympans" (G et D) ont été mesurés sur la TA pour toutes les conditions de l'expérience (cf. section 2.1.5). Les niveaux SPL (G et D) mesurés ont été normalisés par rapport aux niveaux (G et D) de la référence frontale. Les niveaux relatifs ainsi obtenus à 1 kHz et 5 kHz sont présentés dans le tableau 2.1.

Les ILDs relatifs (par rapport à la source de référence frontale) varient entre 0.1 dB et 7.8 dB à 1 kHz et entre 0.1 dB et 17 dB à 5 kHz. Les valeurs maximales sont inférieures de 2 dB (à 1 kHz) et 10 dB (à 5 kHz) aux ILDs mesurés par Sivonen et Ellermeier (2008) sur une tête artificielle B&K (type 4100). En revanche cela correspond aux ILDs estimés à partir de la banque d'HRTFs du Massachusetts Institute of Technology (Gardner *et al.*, 1994) réalisée avec un mannequin KEMAR (modèle DB-4004). D'après ces auteurs, dans le plan horizontal à 60° d'azimut, les ILDs valaient environ 8 dB à 1 kHz et 17 dB à 5 kHz.

DLS médianes :

La figure 2.4 représente les DLS médianes sur l'ensemble des auditeurs ainsi que les écarts interquartiles associés en fonction de l'angle d'incidence. Les niveaux SPL relatifs (normalisés par rapports aux niveaux SPL G et D de la référence frontale) sont également représentés et symbolisés par des triangles bleus.

Comme pour les données individuelles, on constate que l'évolution de la sonie avec l'angle d'azimut varie en fonction de l'angle d'élévation et de la fréquence. La sensibilité en sonie médiane varie avec l'angle d'incidence sur une plage de 3 dB à 1 kHz. A 5 kHz, l'effet de l'angle d'incidence sur la sonie est deux fois plus grand, puisque la sensibilité en sonie évolue sur une plage de 6.5 dB. Cette observation est en parfait accord avec les résultats observés par (Sivonen et Ellermeier, 2006).

Une ANOVA à mesure répétée sur 3 facteurs (Fréq (2) x Az (6) x El (2)) a permis de confirmer les principaux effets observés. L'azimut [$F(4,64)=176.45$; $p<0.001$] et l'élévation [$F(1,16)=69.93$; $p<0.001$] ont un effet très significatif sur la DLS si les facteurs sont considérés indépendamment. Cela montre que l'effet de l'angle d'incidence sur la sonie déjà mis en évidence dans Sivonen et Ellermeier (2006) par exemple, peut être décomposé comme la combinaison des effets de l'angle d'azimut d'une part et de l'angle d'élévation d'autre part.

On relève également une interaction très significative Fréquence x Elévation x Azimut [$F(4,64)=170.43$; $p<0.001$] confirmant le fait que la variation de la DLS avec l'angle d'incidence (Φ, Θ) est fortement dépendante de la fréquence. Si on distingue maintenant les effets de l'azimut et de l'élévation, la variation de sonie avec l'azimut dépend à la fois de la fréquence (variation plus importante à 5 kHz) et de l'élévation (variation plus importante dans le plan horizontal). Cette double dépendance est traduite en termes statistiques par des interactions très significatives : Azimut x Fréquence [$F(4,64)=229.9343$; $p<0.001$] et Azimut x Elévation [$F(4,64)=122.65$; $p<0.001$]. De la même manière, l'évolution de la sonie en fonction de l'élévation dépend à la fois de la fréquence [$F(1,16)=398.7443$; $p<0.001$] et de l'azimut.

Concernant les écarts interquartiles, à 1 kHz ils sont tous inférieurs à 1 dB excepté pour les angles d'incidence ($60^\circ, 0^\circ$), ($90^\circ, 0^\circ$) et ($60^\circ, 45^\circ$). C'est donc pour ces trois incidences que la dispersion entre les auditeurs est la plus grande. A 5 kHz, les écarts interquartiles sont plus grands qu'à 1 kHz (compris entre 1 et 1.6 dB), ce qui révèle une augmentation des différences entre les auditeurs avec la fréquence. Une ANOVA à mesure répétée a permis de mettre en avant l'effet très significatif de la fréquence sur la variabilité interindividuelle [$F(1,13)=25.02$; $p<0.001$]. Cela rejoint les observations déjà faites pour les résultats individuels.

DLS médianes et niveaux SPL :

Pour la majorité des conditions, les DLS sont comprises entre les niveaux SPL relatifs G et D. La relation entre DLS et niveaux SPL est étroite, et d'une manière générale, le niveau SPL de l'oreille ipsilatérale a d'avantage de poids dans le jugement de sonie des différents auditeurs. Dans le plan sagittal, les niveaux SPL relatifs G et D coïncident, les auditeurs comparent alors deux sources en situation diotique. Conformément à ce qui était attendu, le niveau d'égalsation au PSE correspond

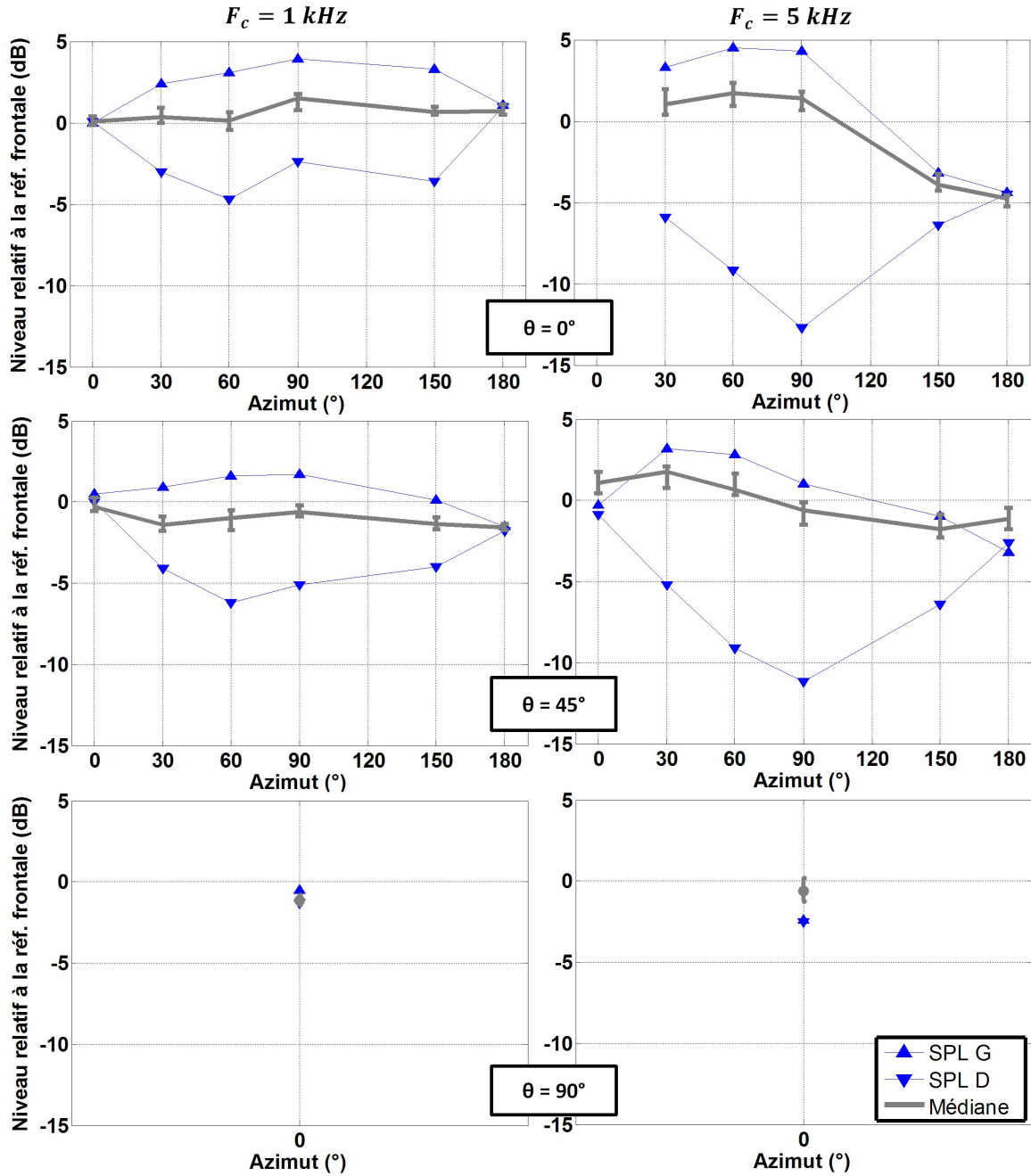


FIGURE 2.4: DLS médianes et écarts interquartiles sur l'ensemble des auditeurs en fonction de l'azimut. Résultats pour les deux fréquences centrales $F_c = 1 \text{ kHz}$ (à gauche) et $F_c = 5 \text{ kHz}$ (à droite) et pour les trois élévations $\Theta = 0^\circ$ (en haut), $\Theta = 45^\circ$ (au milieu) et $\Theta = 90^\circ$ (en bas). Les triangles bleus représentent les niveaux de pression relatifs à l'oreille G (pointe vers le haut) et à l'oreille D (pointe vers le bas).

alors approximativement à la différence de niveau entre les niveaux SPL du son de comparaison et les niveaux SPL de la référence frontale.

Cependant, il y a des exceptions étonnantes à 5 kHz. Les DLS médianes obtenues dans le plan sagittal à élévation non-nulle sont toujours supérieures aux niveaux SPL G et D (écart de l'ordre du décibel), les niveaux SPL relatifs n'étant pas même inclus dans les écarts interquartiles. Ce décalage semble se répercuter également pour l'azimut 150°, proche du plan sagittal. En effet, pour l'incidence (150°, 45°), cinq des auditeurs ont également obtenu une DLS supérieure au niveau SPL G. Il est possible que cela corresponde à un biais lié à la procédure qui opèrerait uniquement pour les élévations non-nulles dans le plan sagittal. Cependant aucune explication pertinente n'a pu être avancée pour le moment.

2.3 Modélisation

En admettant qu'il existe une relation directe entre la sonie globale et les niveaux SPL pour chaque fréquence et que cette relation est donc indépendante de la nature du champ acoustique (cf. § 1.3.3), cette partie consiste à modéliser le plus fidèlement possible cette relation au regard des données expérimentales recueillies. L'approche et le modèle utilisés sont issus des recherches de Robinson et Whittle (1960) et Sivonen et Ellermeier (2008). Différents gains de sommation binaurale de la littérature seront ensuite discutés.

2.3.1 Modélisation des données expérimentales

Suite aux travaux de Robinson et Whittle (1960), Sivonen et Ellermeier (2006), Sivonen (2007) et Sivonen et Ellermeier (2008), deux observations générales peuvent servir de point de départ. D'une part, dans le plan sagittal on est en situation diotique, donc les contributions des niveaux SPL G et D dans la création du percept de sonie binaurale sont les mêmes. D'autre part, lorsque les ILDs sont maximaux, on s'approche d'une situation monaurale et la contribution de l'oreille ipsilatérale devient maximale tandis que la contribution de l'oreille contralatérale tend à s'annuler.

Ainsi, en se basant sur le modèle de Sivonen et Ellermeier (2008), lui-même issu des travaux de Robinson et Whittle (1960), il est envisageable de modéliser la sommation binaurale par une relation non linéaire de la forme suivante :

$$L_{mon_{eq}} = g \times \log_2 \left[2^{L_G/g} + 2^{L_D/g} \right], \quad (2.1)$$

où $L_{mon_{eq}}$ représente le niveau monaural équivalent afin d'obtenir la même sonie que dans une situation diotique ou dichotique caractérisée par les niveaux SPL aux oreilles L_G et L_D . g est défini comme étant le gain binaural (si le même niveau L est présenté aux deux oreilles, le niveau monaural équivalent est $L + g$).

Pour des raisons pratiques, on décidera de définir plutôt le niveau diotique équivalent $L_{diot_{eq}}$ tel que :

$$L_{diot_{eq}} = g \times \log_2 \left[2^{L_G/g} + 2^{L_D/g} \right] - g. \quad (2.2)$$

Afin d'ajuster au mieux le modèle de sommation aux données de l'expérience, la valeur optimale pour le paramètre g (gain binaural) a été recherchée. Cette optimisation est réalisée par minimisation

de l'écart quadratique moyen (EQM) sur l'ensemble des angles d'incidence, entre la DLS mesurée et la DLS prédite par le niveau diotique équivalent. Pour chaque incidence (Φ, Θ) , la DLS prédite correspond à la différence entre le niveau diotique équivalent du son de comparaison ($L_{diot,comp}$) et le niveau diotique équivalent du son de référence ($L_{diot,ref}$). Les valeurs de g parcourues pour la minimisation sont comprises entre 0.01 dB et 99.99 dB avec une résolution de 0.01 dB.

$$EQM = \frac{1}{n} \sum_{(\Phi, \Theta)} \left(DLS - \left[L_{diot,comp(\Phi, \Theta)} - L_{diot,ref} \right] \right)^2 \quad (2.3)$$

Les gains binauraux ont été estimés d'abord de manière individuelle puis sur les données médianes. Dans les deux cas, le gain optimal a été calculé pour chaque fréquence centrale indépendamment puis pour l'ensemble des données (les deux fréquences confondues). Le tableau 2.2 répertorie les différents gains binauraux estimés ainsi que les valeurs de l'EQM associées.

Selon la règle de sommation binaurale proposée dans l'équation 2.2 pour évaluer le niveau diotique équivalent à partir des niveaux SPL, une valeur de g élevée traduit l'attribution d'un poids plus important à l'oreille recevant le plus faible niveau (contralatérale) dans la sommation. Pour un ILD de 20 dB (valeur extrême dans le cadre de notre expérience), une valeur de g de 3 dB correspond à un niveau diotique équivalent inférieur d'environ 3 dB au niveau SPL le plus grand et une valeur de g de 100 dB à un niveau diotique équivalent inférieur d'environ 10 dB au niveau SPL le plus grand. Pour des ILDs "petits" (réalistes), la valeur de g a ainsi peu d'influence sur la DLS estimée à partir des niveaux SPLs (comme le précisaient déjà Robinson et Whittle (1960)). Il arrive que la méthode des moindres carrés aboutisse à des valeurs de g très grands, comme pour l'auditeur TL à 1 kHz par exemple, pour lequel la fonction n'admet pas de minimum local (cf. figure 2.5a). Cependant, d'un point de vue physique, il n'y a pas de sens à considérer un gain binaural supérieur à 10 dB (correspondant à un doublement de sonie à 1 kHz) lors du passage d'une situation monaurale à une situation diotique d'écoute.

Si on considère indépendamment chaque fréquence, on constate que tous les gains sont compris entre 0 dB et 10 dB, à l'exception de 4 "grandes valeurs" à 1 kHz pour les auditeurs TL, RS, WA et MY. A 1 kHz, toutes les valeurs de l'EQM sont inférieures à 0.4. Cela traduit donc finalement un bon ajustement du modèle à cette fréquence. A 5 kHz, les valeurs du gain binaural sont un peu plus stables, étant toutes comprises entre 0.37 dB et 8.71 dB. Pour tous les auditeurs, l'évolution de l'EQM en fonction des valeurs de g possède également un minimum plus précis, reflet de cette stabilité (ex. pour les auditeurs TL et CB, figure 2.5b à 5 kHz). Cependant, à cette même fréquence, les valeurs du minimum de l'EQM sont plus grandes qu'à 1 kHz pour tous les auditeurs. L'explication de l'écart du modèle à 5 kHz se trouve vraisemblablement dans les écarts déjà constatés entre les DLS mesurées et les valeurs attendues, dans le plan sagittal pour les élévations non-nulles (cf. section 2.2.3, p.54). Notons que dans la mesure où ce décalage apparaît principalement dans le plan sagittal ou pour des azimuts proches de 0 ou 180° (soit là où les différences interaurales sont faibles), il devrait avoir peu d'influence dans la détermination de la valeur du gain binaural.

Le gain de sommation binaurale g varie de manière importante entre les auditeurs à 1 kHz et à 5 kHz. Dans une moindre mesure, les valeurs peuvent également varier pour un même auditeur

Auditeurs	Groupe	1 kHz		5 kHz		1 et 5 kHz	
		g	EQM	g	EQM	g	EQM
EP	A	9.14	0.24	8.71	0.77	8.74	0.49
LC	A	3.95	0.37	6.06	0.56	5.76	0.49
TL	A	73.99	0.14	4.68	0.66	5.25	0.56
WA	A	19.52	0.10	4.08	1.08	4.60	0.71
ML	A	6.00	0.09	4.30	2.26	4.48	1.14
CB	A	4.53	0.07	4.46	1.37	4.47	0.69
MY	A	14.56	0.09	3.47	1.21	4.00	0.80
CM	B	3.57	0.15	3.53	0.91	3.54	0.52
SB	B	5.24	0.12	2.94	0.33	3.25	0.28
SL	B	3.68	0.08	3.06	0.33	3.18	0.21
LB	B	1.54	0.37	3.44	0.47	2.89	0.59
MP	B	3.25	0.11	2.75	2.45	2.85	1.24
LG	B	1.95	0.08	3.10	2.31	2.77	1.21
FA	B	3.69	0.21	1.95	0.59	2.33	0.48
MV	B	2.18	0.14	2.07	2.17	2.11	1.11
RS	B	99.99	0.29	0.37	1.83	2.05	2.30
CS	B	3.19	0.20	1.33	1.80	1.89	1.11
Médianes		3.95		3.44		3.25	
Sur les DLS médianes		4.59	0.06	3.31	0.92	3.5	0.49

Tableau 2.2: Valeurs du gain binaural g optimal (eq. 2.2) obtenues par minimisation de l'EQM et valeurs de l'EQM associées. Les deux doubles-colonnes de gauche correspondent à la meilleure estimation pour chaque fréquence centrale et pour chaque auditeur. La double-colonne de droite correspond à la meilleure estimation, toutes fréquences confondues. Les deux dernières lignes représentent respectivement, la médiane des gains binauraux individuels et le gain binaural optimal estimé à partir des DLS médianes.

d'une fréquence à l'autre (jusqu'à environ 2 dB en excluant les grandes valeurs). Si le modèle est ajusté sur l'ensemble des conditions (toutes fréquences confondues), les valeurs de g sont alors légèrement moins dispersées mais continuent de varier entre 2 dB et 9 dB sur l'ensemble des auditeurs. Les valeurs de l'EQM restent également souvent supérieures à 1, pour les mêmes raisons que celles évoquées à 5 kHz.

Finalement, la règle de sommation binaurale varie entre les auditeurs. Les différences interin-

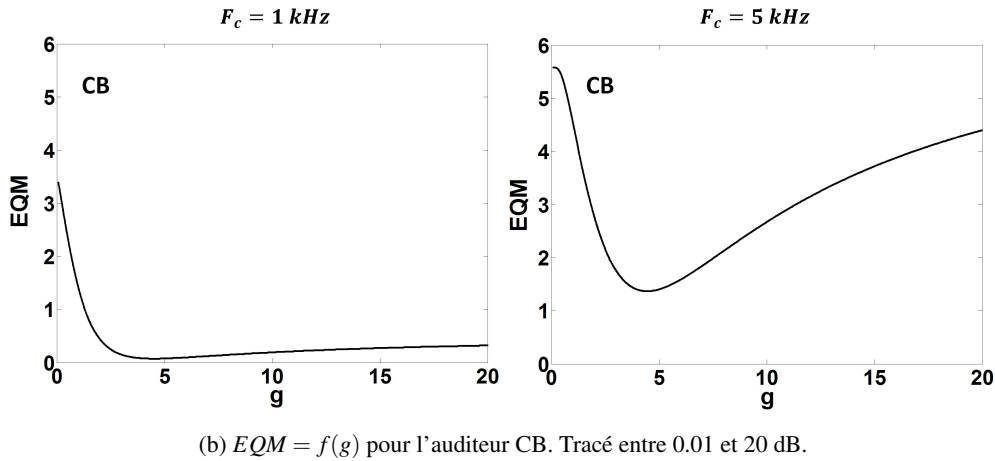
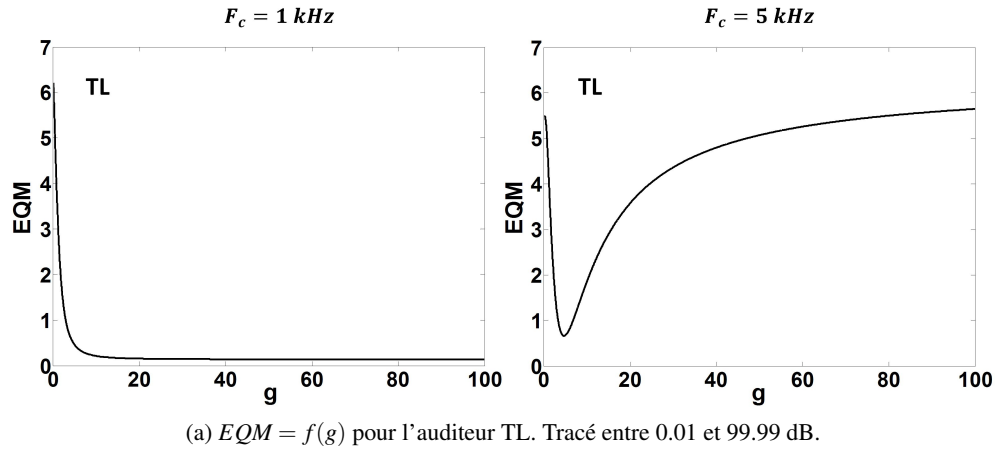


FIGURE 2.5: Écart quadratique moyen en fonction de la valeur du gain binaural pour deux auditeurs (TL et CB) à chaque fréquence centrale (1 kHz à G et 5 kHz à D).

dividuelles observées sur les courbes de DLS individuelles (figure 2.1) sont traduites en terme de variabilité du gain de sommation binaurale sur une plage d'environ 7 dB. En terme de distance par rapport au modèle, la valeur de l'EQM fluctue également beaucoup, pouvant être multipliée par 8 d'un auditeur à l'autre.

D'autre part, l'organisation des gains binauraux et des valeurs d'EQM individuels (obtenus sur l'ensemble des résultats) semblent en accord avec la classification des auditeurs proposée plus haut (cf. § 2.2.2). En effet, les DLS médianes du groupe A étaient toujours plus faibles que les DLS médianes du groupe B. Or les gains binauraux correspondant aux auditeurs du groupe A (moyenne : 4.7 dB) sont toujours plus grands que ceux des auditeurs du groupe B (moyenne : 2.4 dB). Inversement, l'EQM est globalement plus faible pour les auditeurs du groupe A (moyenne : 0.59) que pour les auditeurs du groupe B (moyenne : 1.15). Cela montre donc que les différences de stratégie de sommation binaurale suggérées par la classification des auditeurs se traduisent finalement par des valeurs du gain de sommation binaurale différentes. De plus, d'après les valeurs de l'EQM, le modèle de sommation (équ. 2.1) permet de mieux représenter les stratégies de sommation pour les auditeurs

du groupe A que pour les auditeurs du groupe B.

Enfin, concernant l'estimation sur les données médianes (dernière ligne du tableau 2.2), le gain binaural vaut 4.6 dB à 1 kHz et 3.3 dB à 5 kHz, soit une différence de 1.3 dB entre les deux fréquences. La valeur estimée sur l'ensemble des conditions $g = 3.5$ dB est très proche de la médiane des gains individuels mais également de la valeur estimée à 5 kHz (là où les différences interaurales étaient les plus marquées). D'une manière générale, c'est donc la valeur du gain binaural $g = 3.5$ dB qui permet de prédire au mieux les données issues de cette expérience. C'est également la valeur obtenue par Sivonen et Ellermeier (2008) en utilisant des HRTFs génériques pour un groupe d'auditeurs "experts". Pour les auditeurs "naïfs" la valeur du gain était plus élevée mais la différence entre les deux groupes d'auditeurs n'était pas significative. Le gain de 3 dB finalement retenu par les auteurs (en accord avec d'autres données obtenues en individualisant les HRTFs) reste néanmoins très proche du gain final que nous obtenons.

2.3.2 Comparaison de différents gains de sommation binaurale

Dans la littérature, plusieurs modèles de sommation binaurale ont été proposés. Concernant le modèle initié par Robinson et Whittle (1960) puis repris par Sivonen et Ellermeier (2008), les gains de sommation binauraux suggérés par les auteurs étaient respectivement $g = 6$ dB et $g = 3$ dB. La valeur 6 dB correspond aussi approximativement au gain binaural du modèle de Moore et Glasberg (2007) si on considère une fonction de sonie avec un exposant $\alpha = 0.6$. Concernant la modélisation de nos données, la valeur du gain binaural la plus appropriée a été estimée à 3.5 dB, soit finalement une valeur très proche de 3 dB. Enfin, nous proposons également de confronter les résultats à un modèle plus classique, celui de l'ANSI-S3.4 (2007). C'est un modèle de sommation de sonie parfaite, qui par analogie, peut être représenté par une règle de sommation à 10 dB en faisant l'approximation que la largeur de bande des stimuli considérés est inférieure à la largeur d'une bande critique et que les fonctions de sonie sont des fonctions de puissance ($\alpha = 0.6$). Ces approximations semblent raisonnables dans le cas des stimuli utilisés dans cette expérience.

Afin de comparer nos données expérimentales aux différents gains de sommation binaurale introduits ci-dessus, les DLS médianes sont à nouveau représentées dans la figure 2.6, en parallèle avec les différentes prédictions associées. Toutes les prédictions correspondent à un niveau diotique équivalent calculé à partir des niveaux SPL relatifs G et D.

Nécessairement, les modèles de sommation à 3 et 3.5 dB sont ceux qui permettent de prédire au mieux les données médianes de cette expérience. On constate qu'à 1 kHz, les deux autres modèles fournissent une approximation relativement précise des résultats également. A 5 kHz, dans le plan horizontal, lorsque les ILDs sont grands, le modèle à 6 dB a tendance à sous-estimer légèrement la sonie. Le modèle à 10 dB quant à lui sous-estime la sonie de manière plus importante à 30, 60 et 90° d'azimut. Pour les élévations non nulles, la comparaison avec les différents modèles ne fait que confirmer à nouveau l'existence d'un biais dans les résultats dans le plan sagittal. En effet, pour tous les angles d'azimut 0° ou 180°, les modèles à 3 et 3.5 dB sous-estiment les DLS obtenues de manière systématique. Les modèles à 6 et 10 dB sous-estiment quant à eux la sonie pour tous les angles d'incidence sans distinction.

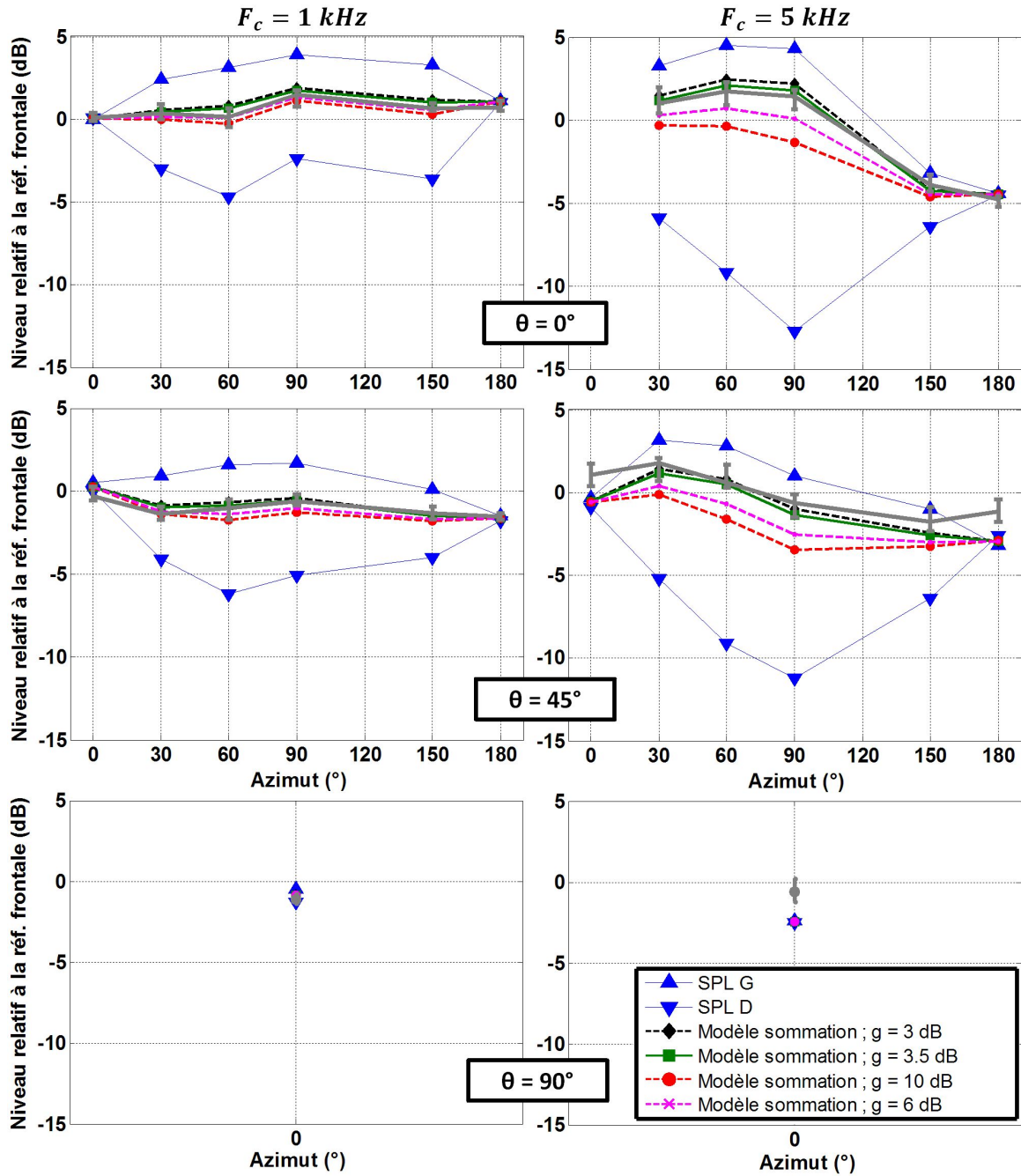


FIGURE 2.6: DLS médianes et écarts interquartiles sur l'ensemble des auditeurs en fonction de l'azimut. Les triangles bleus représentent les niveaux de pression SPLs relatifs à G (pointe vers le haut) et à D (pointe vers le bas). Les courbes noire, verte, violette et rouge représentent respectivement les prédictions des modèles de sommation à 3 dB, 3.5 dB, 6 dB et 10 dB.

Pour conclure, le modèle de Sivonen et Ellermeier (2008) avec un gain binaural de 3 ou 3.5 dB permet de prédire de manière très satisfaisante les données médianes de cette expérience. A 5 kHz, lorsque les ILDs sont grands, le gain individuel optimal variait entre 2 et 6 dB pour l'ensemble des auditeurs (sauf EP). C'est pourquoi le modèle de sommation à 6 dB peut fournir également des prédictions raisonnables dans certains cas. Cependant, le gain binaural de 10 dB correspondant à la règle de sommation binaurale de l'ANSI-S3.4 (2007) sous-estime la sonie dans toutes les situations d'écoute mettant en jeu des grandes différences interaurales.

2.4 Discussion

Les résultats obtenus sont en accord avec les observations des précédentes études sur la sonie de sources spatialisées. Dans les premières investigations de Robinson et Whittle (1960), les DLS moyennes (sur 16 à 20 auditeurs) obtenues pour des bruits centrés à 1.6, 2.5 et 4 kHz étaient comprises entre -6 et +5 dB dans le plan horizontal et entre -5 et 0 dB dans le plan sagittal (pour les élévations positives uniquement). Dans notre étude, ces plages de variations étaient respectivement comprises entre -4.7 et 1.7 dB dans le plan horizontal et entre -4.75 et 1.1 dB dans le plan sagittal. L'effet de l'angle d'azimut sur la sonie était donc légèrement plus prononcé dans l'étude de Robinson et Whittle (1960), mais il ne faut pas oublier que les résultats ne sont pas comparés strictement aux mêmes fréquences.

Dans l'expérience plus récente menée par Sivonen et Ellermeier (2008) en restituant les stimuli au casque, les résultats ont été obtenus pour les mêmes fréquences centrales que dans notre étude (1 kHz et 5 kHz), mais uniquement dans le plan horizontal. A 1 kHz, les DLS moyennes variaient sur une plage d'environ 2 dB (comprise entre 0 et +2 dB) pour le groupe d'auditeurs "naïfs". Dans notre cas, les DLS médianes étaient comprises entre 0 et 1.5 dB pour les mêmes conditions. De la même manière, à 5 kHz les DLS moyennes (ou médianes) étaient respectivement comprises entre -6 et 1 dB (Sivonen et Ellermeier, 2008) et entre -4.7 et +1.7 dB dans notre étude.

L'effet de l'angle d'incidence sur la sonie conserve donc le même ordre de grandeur d'une étude à l'autre, que l'on s'intéresse à l'effet de l'angle d'azimut ou de l'angle d'élévation. Dans toutes les études il a également été montré que les DLS étaient fortement dépendantes de la fréquence, l'effet de l'angle d'incidence augmentant avec la hauteur de la fréquence. Cette dépendance a été à nouveau mise en avant dans cette expérience entre les deux fréquences centrales 1 et 5 kHz. D'autre part, l'investigation du plan à 45° d'élévation a permis de mettre clairement en évidence l'effet de l'angle d'élévation sur la sonie, permettant de compléter la littérature sur ce point. Finalement, l'ensemble de ces observations est cohérent, dans le sens où elles sont toutes directement liées à la variation du niveau global des signaux parvenant aux oreilles (dans le plan sagittal) ainsi qu'à la variation des différences interaurales (en dehors du plan sagittal). Ces variations sont causées par l'effet d'obstruction de la tête de l'auditeur et des pavillons, dont l'amplitude de variation en fonction de l'angle d'azimut est globalement plus importante en hautes fréquences et pour des élévations faibles (Blauert, 1997; Wersényi, 2010).

Le modèle de sommation binaurale proposé par Sivonen et Ellermeier (2008) a ensuite été ajusté afin de modéliser le mieux possible les DLS obtenues à partir des niveaux SPL "aux tympans". Le gain binaural optimal variait entre 2 et 8.7 dB sur l'ensemble des auditeurs. Dans Sivonen et Eller-

meier (2006) ce gain variait entre 0.7 et 9.1 dB, sachant que les sources étaient des haut-parleurs et que les niveaux SPL étaient estimés à partir des HRTFs individuelles. En ayant recours à des HRTFs génériques, la plage de variation du gain binaural était cependant légèrement réduite entre 1.3 et 8.4 dB (Sivonen et Ellermeier, 2008). Finalement, nous obtenons une variabilité du gain binaural équivalente à celle rapportée par ces auteurs. Concernant l'estimation sur les données médianes, la valeur optimale retenue pour g est 3.5 dB. Cette valeur est très proche du gain de 3 dB proposé par Sivonen et Ellermeier (2008) mais reste bien inférieure à la valeur 6 dB issue des travaux de Robinson et Whittle (1960) ou Moore et Glasberg (2007). En terme de rapport de sonie binaural-à-monastral, un gain de 3 dB correspond approximativement à un rapport de 1.2 (pour $\alpha = 0.6$). Cette valeur est donc inférieure à la plupart des valeurs issues des études évaluant le rapport de sonie binaural-à-monastral pour des stimuli non-spatialisés au casque (e.g. Marks (1978) $\simeq 2$, Zwicker et Zwicker (1991) $\simeq 1.5$, Marozeau *et al.* (2006) $g \simeq 1.3$).

Enfin, l'analyse des résultats de manière individuelle a permis de mettre en avant l'importance des différences entre les auditeurs, d'abord du point de vue des sensibilités directionnelles en sonie puis dans les valeurs du gain de sommation binaurale estimé à partir des données. Une classification ascendante hiérarchique des auditeurs à partir des DLS a permis de mettre en évidence le lien entre la variabilité observée dans les DLS et la variabilité observée dans la valeur du gain binaural, reflétant des stratégies de sommation binaurale différentes. Cependant, l'origine de cette variabilité interindividuelle n'est pas claire.

En présence d'une source réelle, une part de la variabilité dans la DLS entre les auditeurs est comprise dans la variabilité des HRTFs (Sivonen et Ellermeier, 2006). Mais les réponses impulsionnelles utilisées ici pour synthétiser les stimuli étaient génériques, ce qui implique que les mêmes signaux ont été présentés aux oreilles de tous les auditeurs. Les réponses impulsionnelles du casque étaient également génériques, ne tenant pas compte des particularités du couplage du casque sur la tête de chaque auditeur. Cela a pu induire des différences de niveaux interindividuelles pour une direction donnée. Mais ces différences ne sont pas dépendantes de la direction et ont donc été prises en compte lors de la comparaison avec l'incidence frontale. Malgré la procédure permettant de s'affranchir des caractéristiques physiques individuelles, la variabilité interindividuelle est donc encore bien présente.

Ces différences entre les auditeurs dans la formation du percept de sonie globale pourraient trouver leur origine à différents niveaux. Elles pourraient être causées par des différences de sensibilités monaurales, des différences dans le processus de sommation binaurale au niveau central, par l'implication d'autres indices binauraux comme les ITDs ou encore par l'altération plus ou moins importante de la sonie par certains facteurs cognitifs.

Une des hypothèses avancées par Sivonen et Ellermeier (2006) est basée sur le phénomène de constance perceptuelle en sonie ("loudness constancy") qui décrit la constance d'un percept en dépit de la variation de certains paramètres physiques (Zahorik et Wightman, 2001). Dans le cas de la distance, le percept est la sonie et le paramètre physique est la distance de la source. Ce concept pourrait être étendu à la situation où c'est l'angle d'incidence de la source qui varie (à distance égale), se traduisant par des variations dans les signaux parvenant aux oreilles de l'auditeur. Ainsi, selon l'attitude adoptée par l'auditeur (jugement de "la puissance de la source" ou bien de "l'énergie qui en résulte"), la variation des différences interaurales avec l'angle d'incidence aurait une influence plus ou moins grande sur la sensibilité directionnelle en sonie. Autrement dit, si l'auditeur est capable d'estimer la

puissance de la source, le jugement de sonie pourrait-être biaisé par l'idée qu'il s'en fait. L'apparition d'un tel biais dans la sonie directionnelle implique d'une part une bonne localisation de la source (ce qui semble être le cas dans le plan horizontal), mais serait probablement favorisée par l'existence d'un contexte plus marqué (indices dynamiques liés au mouvement de la source, réverbération,...). En effet, la constance perceptuelle en sonie a été mise en évidence dans une acoustique de salle réelle (Zahorik et Wightman, 2001), impliquant l'existence d'un contexte fixe fournissant des indices importants sur la distance de la source et sa puissance (rapport d'énergie entre le son direct et le son réverbéré par exemple). Dans le cadre de notre expérience le contexte était peu défini (stimuli artificiels, environnement anéchoïque, écoute au casque), ce qui ne tend pas à renforcer cette hypothèse. Il serait toutefois intéressant de quantifier l'apparition potentielle du phénomène dans un tel contexte en mélangeant plusieurs sources de puissance différentes suivant différents angles d'incidence et en demandant aux auditeurs de juger la puissance de la source (consigne "objective").

Une autre explication plus basique, pourrait-être l'implication de la variabilité interindividuelle des fonctions de sonie dans la variabilité des DLS. En effet, un certain nombre d'études rapportent des différences importantes dans les fonctions de sonie monaurales individuelles (e.g. Hellman et Meiselman (1988), Stevens et Guirao (1964)). Ces différences sensorielles pourraient intervenir avant le processus de sommation binaural et se répercuter sur le percept de sonie globale. Un protocole expérimental permettant de tester cette hypothèse est ainsi proposé dans le chapitre suivant.

2.5 Conclusion

L'étude sur la sonie directionnelle pour une source de bruit à bande étroite proposée par Sivonen et Ellermeier (2008) a été reproduite dans un nouveau contexte expérimental (laboratoire, matériel, auditeurs,...) sur un plus grand nombre d'auditeurs et pour une plus grande variété d'angles d'incidence. Nos principales conclusions concernant les sensibilités en sonie directionnelles sont en accord avec celles de V.P. Sivonen et W. Ellermeier : la sonie varie de manière importante en fonction de l'angle d'incidence (jusqu'à 9 dB pour les données individuelles, jusqu'à 6.5 dB pour les données médianes), cette variation est fortement dépendante de la fréquence (l'effet de la direction étant plus important à 5 kHz qu'à 1 kHz) et il existe des grandes différences interindividuelles de sensibilité directionnelle en sonie. Quelques éléments nouveaux ont également pu être avancés.

Les mesures avec des angles d'azimut variables dans le plan à 45° d'élévation ont permis de mettre en évidence l'effet général de l'angle d'élévation sur la sonie (tous azimuts et fréquences confondus). Cela se traduit par une variation plus importante de la sonie dans le plan horizontal que dans le plan à 45° d'élévation.

De plus, la participation d'un plus grand nombre d'auditeurs à l'expérience (17) a permis de confirmer l'amplitude de la variabilité entre les auditeurs déjà observée par Sivonen et Ellermeier (2008). Une classification ascendante hiérarchique des auditeurs a suggéré la formation de deux groupes se distinguant par des niveaux de sensibilité en sonie différents (niveau "fort" ou "faible"). Malgré des différences faibles entre les groupes, l'orientation de chaque groupe était étonnamment stable sur l'ensemble des angles d'incidence. Un lien a finalement pu être établi entre ces groupes et le gain binaural, reflet de stratégies individuelles de sommation binaurale différentes.

Le modèle proposé par Sivonen et Ellermeier (2008) a été ajusté à partir des données expérimentales. Un gain de sommation binaurale de 3.5 dB a permis de représenter avec le plus de précision

possible les données médianes recueillies. La réplication de l'expérience menée par Sivonen et Ellermeier (2008) a finalement permis de confirmer les résultats rapportés par ces auteurs et donc de s'accorder à conclure qu'un gain binaural d'environ 3 dB permet de prédire avec précision la sonie d'une source artificielle spatialisée dans un environnement réel.

Enfin, une hypothèse sur l'implication de la variabilité des fonctions de sonie monaurales individuelles dans la variabilité des DLS a été avancée et constitue l'objet du chapitre suivant.

Influence de la variabilité des fonctions de sonie monaurales individuelles

L'objectif de ce chapitre est de tester l'hypothèse selon laquelle une partie de la variabilité interindividuelle constatée dans les égalisations en sonie (chapitre 2) peut être expliquée par des différences d'exposants entre les fonctions de sonie des auditeurs. Dans ce but les fonctions de sonie monaurales ont été estimées pour quatre auditeurs à l'aide de deux méthodes de mesure : une estimation de grandeur (*ME*) et une production de rapport (*RP*).

3.1 Mesure des fonctions de sonie par estimation de grandeur (*ME*)

3.1.1 Méthode

3.1.1.1 Auditeurs

Quatre des dix-sept auditeurs ayant participé aux égalisations en sonie (chapitre 2) ont été sollicités pour participer à cette expérience. Le choix des auditeurs (EP, RS, FA, et MV) a été réalisé sur la base des sensibilités directionnelles en sonie mesurées. Afin de pouvoir mettre en évidence l'implication des exposants des fonctions de sonie dans les différences de DLS, nous avons choisi des auditeurs ayant à la base des sensibilités directionnelles en sonie bien contrastées. Ainsi, les auditeurs EP et RS ont été retenus pour avoir obtenu des résultats situés aux extrêmes à 5 kHz (chapitre 2, figure 2.1), fréquence pour laquelle les différences interaurales étaient les plus grandes. Ce n'était cependant pas le cas à 1 kHz. FA et MV avaient quant à eux obtenu des résultats intermédiaires et ont donc été choisis pour représenter des tendances plus proches de la médiane.

Les quatre auditeurs étaient des hommes âgés entre 26 et 50 ans. Aucune perte auditive n'a été relevée pour les auditeurs FA, EP et MV (seuils < 20 dB HL). On notera en revanche une perte légère à l'oreille gauche de l'auditeur RS, pour laquelle un seuil d'audition de 30 dB HL a été mesuré à 1 kHz. Le détail des audiogrammes est présenté dans l'annexe A.

3.1.1.2 Stimuli

Les stimuli utilisés étaient des bruits roses filtrés par bandes de tiers d'octave centrées à 1 kHz et 5 kHz (identiques aux bruits utilisés pour simuler les sources spatialisées dans le chapitre 2). Les sons avaient une durée de 1 s, incluant une croissance et une décroissance en cosinus de 20 ms.

Sept niveaux de présentation différents ont été considérés, s'échelonnant de 50 à 80 dB SPL par pas de 5 dB. Cela permettait d'englober l'ensemble de la plage des niveaux SPL impliqués dans la première expérience de sonie directionnelle (chapitre 2). En effet, dans cette dernière les niveaux SPL

(G et D) étaient tous compris entre 53 et 79 dB SPL, en prenant en compte les niveaux d'égalisation au PSE pour l'ensemble des auditeurs et des conditions expérimentales.

Le dispositif expérimental était strictement identique à celui de l'expérience précédente. Les sons étaient générés à l'aide d'une carte son Echo Gina 24 (16 bits, 44100 Hz) puis restitués à l'auditeur en monaural à l'aide d'un casque Sennheiser HD 650 dans une cabine audiométrique.

3.1.1.3 Procédure

Les exposants des fonctions de sonie monaurales individuelles ont été estimés à l'aide d'une méthode d'estimation de grandeur sans référence (e.g. Logue (1976), Collins et Gescheider (1989), Hellman et Meiselman (1993)). Il était demandé à chaque auditeur d'assigner librement un nombre dont la valeur représentait au mieux la sensation de niveau sonore induite par chaque stimulus. Aucune valeur de référence n'était fournie. Avant chaque série de mesures, la consigne était rappelée et formulée de la manière suivante : *"Vous allez entendre une série de sons de différents niveaux. Votre tâche est d'évaluer le niveau sonore de chacun des sons en leur assignant des nombres. Vous êtes libres d'utiliser tout nombre positif qui vous semble approprié (entier, décimal,...). Le nombre zéro est réservé au cas où vous n'entendriez aucun son. Ne tenez pas compte de la valeur assignée au son précédent, contentez-vous de choisir la valeur la plus appropriée pour le son que vous serez en train de juger. Une fois votre jugement effectué, validez votre choix pour jouer le son suivant."* Après chaque présentation, l'auditeur était invité à entrer et valider sa réponse. Un intervalle de temps fixe de 500 ms était inséré avant la présentation du stimulus suivant.

Les 4 conditions expérimentales (2 fréquences \times 2 oreilles) ont été réparties en 4 blocs distincts et l'ordre des blocs était soumis à un tirage aléatoire pour chaque auditeur. Les 4 blocs étaient réalisés à la suite (dans la même heure) afin d'éviter que les auditeurs modifient leur échelle de nombre d'un bloc à un autre, suite à la modification d'un paramètre (fréquence ou oreille de présentation).

L'ordre de présentation des différents niveaux dans chaque bloc était organisé en accord avec l'algorithme proposé par Cross (1973). En effet, plusieurs études ont montré que pour une séquence de stimuli, la réponse fournie pour le stimulus de rang n était corrélée positivement avec la réponse fournie pour le stimulus précédent de rang $n - 1$ (e.g. Cross (1973), Ward (1973)). Cette tendance de l'auditeur à répéter sa réponse précédente est connue sous le nom "d'effet d'assimilation" et se traduit dans l'estimation de grandeur par une augmentation du jugement de sonie lorsque la réponse précédente était grande et une diminution du jugement de sonie lorsque la réponse précédente était faible. Or, les niveaux faibles étant statistiquement plus souvent présentés après un niveau plus élevé (et inversement), l'effet d'assimilation contribuerait finalement à sous-estimer l'exposant des fonctions de sonie. Ainsi, afin d'évaluer la quantité "d'assimilation" et maîtriser les biais séquentiels, il est suggéré dans Cross (1973) d'organiser la séquence des différents stimuli pour que chaque niveau soit présenté une fois (et une fois seulement) après chacun des différents niveaux testés (le niveau considéré inclus). Toutes les configurations possibles de succession de deux niveaux ont ainsi été représentées. Chaque niveau a donc été répété N fois (si N est le nombre de niveaux) et un stimulus supplémentaire était nécessairement ajouté en position initiale pour commencer la série. Cela faisait un total de $N \times N + 1$ présentations, soit 50 présentations par bloc pour 7 niveaux différents. Pour chaque nouveau bloc, un nouveau tirage pseudo-aléatoire de l'ordre des 50 stimuli à présenter était ainsi effectué, en respectant l'algorithme. Un total général de 200 estimations (4×50) ont été effectuées par chaque auditeur.

L'expérience a été répétée deux fois dans sa globalité dans deux sessions différentes, séparées par un intervalle de temps variant entre 11 et 20 jours selon les auditeurs (EP et FA : 11 j, MV : 16 j, RS : 20 j).

Des blocs d'entraînement ont également été mis en place au préalable de chaque session afin de familiariser les auditeurs avec les niveaux présentés (2 blocs pour la session 1 et 1 bloc pour la session 2). Les conditions expérimentales ainsi que l'ordre de la séquence de niveaux étaient tirés de manière aléatoire pour chaque bloc d'entraînement. Les résultats issus des sessions d'entraînement n'ont pas été pris en compte dans l'analyse des résultats du test.

3.1.2 Résultats

3.1.2.1 Fonctions de sonie, méthode de Cross

Le modèle de fonction de sonie utilisé pour approcher les données est le modèle le plus simple, initialement proposé par Stevens (1955) et reliant l'intensité (ou la pression acoustique P) à la sensation (S) par une simple relation de puissance de la forme $S = k.P^\alpha$. Cependant, l'analyse étant portée sur l'influence d'une différence de niveau sur la sonie et non sur la correspondance exacte entre le niveau de pression et la sonie, seules les valeurs du coefficient directeur α des fonctions de sonie ont été estimées.

Selon l'approche de Cross (1973), l'effet d'assimilation se traduit par un biais séquentiel systématique induisant une surestimation de la sonie du stimulus i si le niveau du stimulus précédent j est plus grand et inversement, une sous-estimation de la sonie de i si le niveau de j est plus faible. Ainsi, l'estimation numérique N_{ij} fournie par les auditeurs en réponse au stimulus i est proportionnelle à la fonction de puissance "réelle" reliant la sensation à l'intensité du stimulus, le biais dépendant du niveau du stimulus précédent : $N_{ij} = k \times (P_i)^\alpha \times \left(\frac{P_j}{P_i}\right)^b$.

Soit, $N_{ij} = k \times (P_i)^m \times (P_j)^b$ où $m = \alpha - b$.

En prenant en compte le détail de chacune des 49 estimations effectuées dans chaque bloc, il est donc possible d'estimer m et b à l'aide d'une régression linéaire multiple sur le logarithme des estimations, en fonction du niveau du stimulus et du niveau du stimulus précédent (cf. Cross (1973), éq. (6)) :

$$\log(N_{ij}) = \log(k) + m \times \log(P_i) + b \times \log(P_j), \quad (3.1)$$

Les estimations individuelles obtenues pour chaque fréquence et chaque oreille (moyenne géométrique sur 7 estimations) ainsi que les fonctions de sonie associées (de pente $\alpha = m + b$) sont représentées dans la figure 3.1. La ligne du haut et la ligne du bas représentent respectivement les résultats obtenus lors de la 1^{re} et de la 2^e session.

On constate dans un premier temps que les écart-types (géométriques) correspondant à la variabilité des sept estimations réalisées pour un même niveau sonore, sont plus grands pour les auditeurs EP et RS que pour les auditeurs MV et FA. Cela traduit donc soit une meilleure répétabilité pour les auditeurs MV et FA, soit une plus grande influence du niveau du stimulus précédent (j) pour les

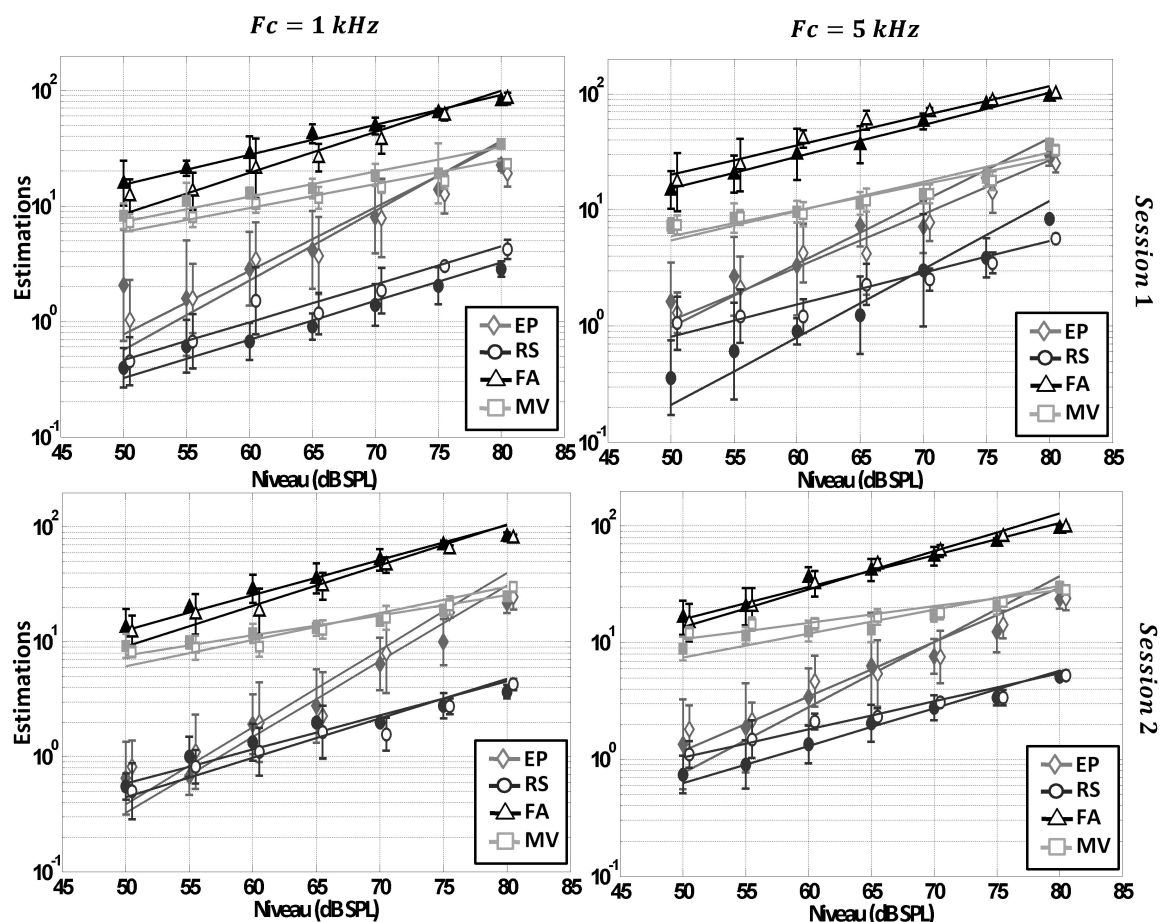


FIGURE 3.1: Fonctions de sonie individuelles à 1 kHz (à gauche) et à 5 kHz (à droite) obtenues pour les 4 auditeurs lors de la session 1 (en haut) et la session 2 (en bas). Les symboles pleins ou vides représentent respectivement les moyennes géométriques des estimations effectuées pour chaque niveau SPL, pour l'oreille G ou pour l'oreille D. Les barres d'erreurs correspondent aux écarts-types associés. Les droites représentent les fonctions de sonie estimées par la méthode de Cross (1973), prenant ainsi en compte l'effet d'assimilation induit par l'organisation séquentielle des niveaux.

auditeurs EP et RS. Pour ces deux auditeurs, la variabilité est d'autant plus importante que le niveau est faible. Cela sous-entendrait donc un effet séquentiel plus prononcé aux faibles niveaux.

D'une manière générale, les fonctions de sonie sont relativement similaires d'une session à l'autre et pour les deux fréquences centrales. En revanche des différences importantes apparaissent entre les différents auditeurs, autant dans les échelles de nombre utilisées (positions relatives des droites) que dans les rapports de nombres utilisés pour juger d'une même différence de niveau (coefficients directeurs). Par ailleurs, on constate que les pentes obtenues pour les deux oreilles sont très proches. Une exception subsiste cependant pour l'auditeur RS à 5 kHz dans la session 1. En effet, les exposants obtenus ont pour valeur 1.17 à l'oreille G et 0.55 à l'oreille D, asymétrie qui sera discutée dans la partie 3.1.2.2.

D'autre part, selon Cross (1973) l'effet d'assimilation amène à sous-estimer l'exposant obtenu par estimation de grandeur si le contexte local n'est pas pris en considération. Les résultats obtenus sont en accord avec cette théorie puisque les exposants estimés en appliquant la méthode de Cross étaient dans tous les cas plus grands que les exposants qui auraient été obtenus en prenant en compte uniquement l'estimation moyenne pour chaque niveau de présentation (sans prendre en compte l'influence des niveaux précédents). C'est pour l'auditeur EP que les effets séquentiels semblaient avoir le plus d'impact, notamment à 1 kHz. A titre d'exemple, pour cet auditeur à 1 kHz dans la session 1 (figure 3.1, en haut à gauche, symboles triangles), les exposants estimés simplement à partir des estimations moyennes pour chaque niveau (régression simple) valaient respectivement 0.78 et 0.85 pour les oreilles G et D (droites suggérées respectivement par les triangles pleins et vides), tandis que les exposants estimés en tenant compte du niveau précédent pour chaque estimation (régression multiple) valaient respectivement 1.10 et 1.21 (droites représentées sur la figure). Sur l'ensemble des auditeurs et des conditions expérimentales, la prise en compte de ce biais séquentiel correspondait à une augmentation de la valeur de l'exposant comprise entre 0.01 (FA, session 1, 5 kHz) et 0.35 (EP, session 1, 1 kHz).

3.1.2.2 Comparaison des sessions de mesure 1 et 2

Deux sessions de mesure ont été réalisées par chaque auditeur à au moins dix jours d'intervalle. Cela a permis entre autre d'évaluer la répétabilité de la méthode pour estimer les exposants des fonctions de sonie.

Les exposants α obtenus pour chaque condition expérimentale dans chacune des sessions ont été comparés deux à deux à l'aide d'un test de Student, afin de tester l'égalité statistique des exposants d'une session à l'autre. Les résultats sont présentés dans le tableau 3.1.

La probabilité p est strictement inférieure à 5% dans 13 cas sur 16. L'hypothèse nulle (stipulant l'égalité des exposants) devra donc être rejetée dans la plupart des cas. Les fonctions de sonie sont donc statistiquement différentes d'une session à l'autre. Il faut cependant modérer l'importance de ces différences entre les exposants. En effet, elles restent toutes comprises entre 0.01 et 0.17, hormis pour l'auditeur EP à 1 kHz à G (0.22) et l'auditeur RS à 5 kHz à G (0.53). Dans ces deux cas de figure, la source de l'écart semble principalement associée à l'estimation des exposants de la session 1 (1.10 pour EP et 1.17 pour RS), dont les écarts-types associés sont les deux plus grands sur l'ensemble des deux sessions (respectivement 0.09 et 0.08).

La différence entre les deux sessions reste cependant la plus grande pour l'auditeur RS à 5 kHz pour l'oreille G. Pour cette condition, l'écart entre les exposants de l'oreille G et de l'oreille D lors de la 1^{re} session était de 0.62, tandis que ce même écart était égal à 0.17 pour la session 2. Il y a donc un écart important entre les exposants estimés pour chaque oreille lors de la 1^{re} session.

Dans la littérature, il est admis que les fonctions de sonie monaurales pour l'oreille G et pour l'oreille D sont identiques (e.g. Marks (1978), Marks (1980)). Cette hypothèse a également été vérifiée par Marozeau et Florentine (2009) pour des auditeurs présentant des pertes auditives similaires entre les deux oreilles (différences de seuils d'audition de moins de 3 dB entre les deux oreilles). En tous cas il n'existe pas, à notre connaissance, d'étude qui mette en évidence des différences entre les

$$1. RC = \frac{|\alpha(S1) - \alpha(S2)|}{\sqrt{\frac{Sc_1^2}{n_1} + \frac{Sc_2^2}{n_2}}} \text{ avec } n_1 = n_2 = 49 \text{ (nombre de valeurs } (P_i; P_j) \text{ des variables explicatives pour la régression).}$$

Auditeur	F_c	Oreille	Session 1		Session 2		$\alpha(S2) - \alpha(S1)$	RC	p
			α	ET	α	ET			
EP	1000	G	1.10	0.09	1.32	0.06	0.22	5.17	< 0.05
EP	1000	D	1.21	0.06	1.34	0.07	0.14	10.16	< 0.05
EP	5000	G	1.08	0.07	1.12	0.07	0.04	2.88	< 0.05
EP	5000	D	0.92	0.06	0.93	0.05	0.01	1.15	—
RS	1000	G	0.67	0.04	0.59	0.03	-0.07	9.79	< 0.05
RS	1000	D	0.66	0.06	0.69	0.05	0.03	3.15	< 0.05
RS	5000	G	1.17	0.08	0.65	0.04	-0.53	43.90	< 0.05
RS	5000	D	0.55	0.05	0.48	0.02	-0.07	9.11	< 0.05
FA	1000	G	0.52	0.03	0.61	0.03	0.09	16.52	< 0.05
FA	1000	D	0.71	0.04	0.71	0.03	-0.01	0.82	—
FA	5000	G	0.55	0.04	0.55	0.03	0.00	0.16	—
FA	5000	D	0.51	0.04	0.65	0.03	0.14	20.54	< 0.05
MV	1000	G	0.43	0.04	0.35	0.02	-0.08	14.48	< 0.05
MV	1000	D	0.41	0.02	0.46	0.02	0.05	12.19	< 0.05
MV	5000	G	0.51	0.03	0.41	0.02	-0.10	18.57	< 0.05
MV	5000	D	0.45	0.03	0.28	0.01	-0.17	37.41	< 0.05

Tableau 3.1: Comparaison des exposants entre la session 1 et la session 2. Sont présentés (de gauche à droite) : les exposants α estimés pour les sessions 1 et 2 ainsi que les écarts types (ET) associés ; la différence des exposants entre les deux sessions ; le rapport critique (RC)¹ ; la probabilité p de rejeter à tort l'égalité des deux exposants.

pentés des fonctions de sonie G et D d'un même auditeur de l'ordre de celles que nous observons ici. Certaines études (par ex. Emmerich *et al.* (1988); Schneider et Cohen (1997)) ont montré un avantage de l'oreille D sur l'oreille G pour une présentation dichotique de stimuli de type « parole » principalement (impliquant spécifiquement l'hémisphère gauche du cerveau dans le traitement de l'information). Mais ce n'est pas le cas ici et l'avantage observé est au contraire pour l'oreille G. L'asymétrie observée ne provient pas non plus d'une différence des seuils d'audition entre les deux oreilles car la seule perte relevée pour cet auditeur a été mesurée à 1 kHz et non à 5 kHz. Cela remet donc en question la confiance que l'on peut accorder aux estimations des coefficients pour cette condition. L'exposant obtenu pour l'oreille D dans la session 1 (0.55) est plus cohérent avec les deux exposants obtenus lors de la 2^e session (0.65 et 0.48) que l'exposant estimé pour l'oreille G (1.17). De plus, l'écart-type associé à l'estimation de l'exposant pour l'oreille G dans la session 1 est plus grand que la moyenne. Dans le cas de l'auditeur RS, le coefficient $\alpha(S1)$ à 5 kHz pour l'oreille G a donc finalement été écarté, seule la valeur $\alpha(S2)$ a été conservée pour cette condition.

D'une manière générale, bien qu'il existe des différences au sens statistique entre les exposants

estimés pour chacune des deux sessions, ces différences restent faibles. De plus, le soin a été pris d’instaurer des séances d’entraînement avant chaque session afin d’éviter les effets d’apprentissage susceptibles d’apparaître au début des sessions. Nous avons donc finalement choisi de réunir les résultats des sessions 1 et 2 sans distinction.

3.1.2.3 Résultats sur l’ensemble des deux sessions

Suite aux précédentes conclusions, les données des deux sessions ont été combinées avant d’effectuer la régression. Ainsi, pour chaque condition (fréquence \times oreille), toutes les estimations N_{ij} obtenues lors de la session 1 ont été moyennées (moyenne géométrique) avec les estimations correspondantes de la session 2. La régression linéaire a ensuite été effectuée sur l’ensemble de ces estimations moyennes. Les exposants α ainsi obtenus sont représentés dans la figure 3.2.

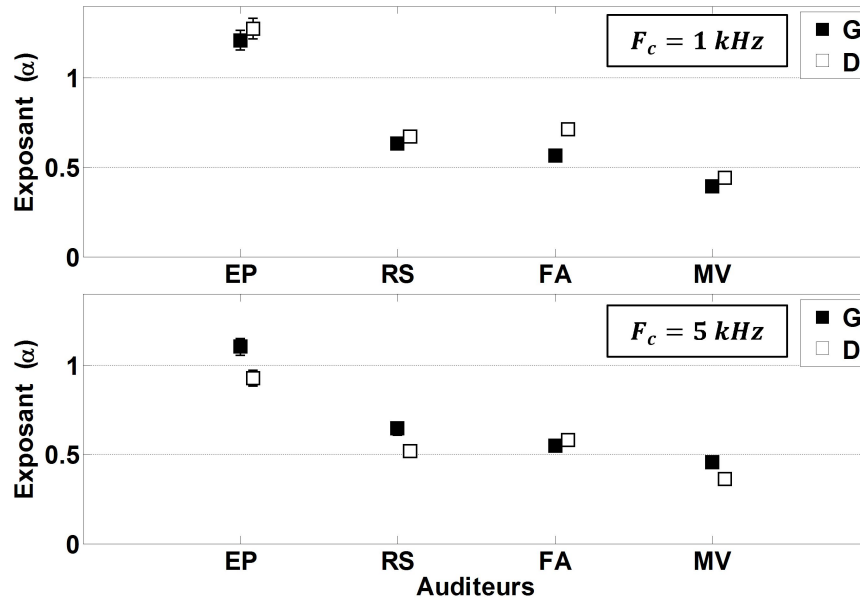


FIGURE 3.2: Exposants obtenus à l’issu du moyennage des données sur les deux sessions, pour chacun des 4 auditeurs. Les symboles pleins représentent les résultats pour l’oreille G et les symboles vides les résultats pour l’oreille D. Les graphiques haut et bas correspondent respectivement aux fréquences centrales 1 kHz et 5 kHz.

On constate que l’ordre de grandeur des exposants pour chaque auditeur est relativement stable sur les deux fréquences centrales et que les différences d’exposant entre les deux oreilles sont généralement faibles. On rappellera que pour l’auditeur RS à 5 kHz, l’exposant à gauche est issu uniquement des données obtenues lors de la session 2, c’est pourquoi la différence entre les deux oreilles pour cet auditeur est également faible à 5 kHz. Cependant il existe des différences interindividuelles importantes, la valeur de l’exposant pouvant varier du simple au triple entre deux auditeurs.

Les exposants estimés pour chaque auditeur et chaque oreille ont été comparés deux à deux (même méthode que pour la comparaison entre les deux sessions), afin de tester l’égalité statistique

d'une fréquence à l'autre. L'analyse a montré que les exposants étaient statistiquement différents pour chaque fréquence dans la majorité des cas (sauf pour EP et FA à l'oreille gauche). De plus, la variation de la valeur de l'exposant en fonction de la fréquence était globalement plus importante pour l'oreille D et variait d'un auditeur à l'autre. Les différences d'exposants minimales et maximales entre les deux fréquences pour l'oreille droite étaient respectivement de 0.07 (MV) et 0.35 (EP). Cependant, à part pour EP à l'oreille droite, toutes les différences d'exposants entre les deux fréquences étaient finalement inférieures à 0.15. Certaines études ont mis en évidence l'influence de la fréquence sur les fonctions de sonie. Les exposants ont en effet tendance à prendre des valeurs plus grandes en-dessous de 400 Hz (Hellman et Zwislocki, 1968; Schneider *et al.*, 1972). Il est d'autre part couramment admis que les fonctions de sonie ont la même pente au-delà de 1 kHz, ce qui va à l'encontre de nos observations. La suite de l'analyse sera tout de même menée pour chaque fréquence indépendamment, distinction souhaitée pour essayer d'expliquer au mieux la variabilité des DLS mesurées à chaque fréquence.

D'autre part, les différences interindividuelles déjà observées dans la figure 3.1 sont à nouveau mises en évidence. Les exposants obtenus sont compris entre 0.36 et 1.27 toutes conditions confondues et entre 0.39 et 1.27 à 1 kHz. Cette plage de variation correspond par exemple aux ordres de grandeur observés par Collins et Gescheider (1989) à l'aide d'une méthode similaire (AME) pour des sons purs à 1 kHz. Des exposants allant de 0.3 à 1.5 sur un panel de 23 auditeurs (adultes et enfants confondus) avaient été rapportés par les auteurs. De la même manière des exposants individuels compris entre 0.44 et 0.78 ont été estimés par Hellman et Meiselman (1993) à l'aide d'une AME pour des sons purs à 500 Hz. La plage de variation était alors plus restreinte, mais cette différence peut être due à la différence de fréquence avec les autres stimuli.

Enfin, concernant la symétrie entre les deux oreilles, les exposants obtenus pour l'oreille G et pour l'oreille D ont également été comparés deux à deux pour chaque auditeur et chaque fréquence. L'égalité des exposants doit être acceptée ($p = 0.05$) pour EP et RS à 1 kHz et pour FA à 5 kHz, mais est rejetée pour les cinq autres conditions. Cela semble traduire une très légère augmentation de la pente des fonctions de sonie pour l'oreille D à 1 kHz et inversement, une diminution de la pente pour l'oreille D à 5 kHz. Cependant ces différences d'exposant entre G et D sont comprises entre 0.03 et 0.18 (pour une moyenne de 0.09), sur l'ensemble des deux fréquences. Elles restent donc très faibles au regard de certaines différences qui existent entre les auditeurs.

Bien que des différences faibles mais significatives aient été observées entre les exposants obtenus pour l'oreille G et pour l'oreille D, nous ne connaissons pas de références mettant en avant une influence de l'oreille de présentation sur la pente des fonctions de sonie monaurales dans les conditions de notre expérience (cf. § 3.1.2.2). Il nous a donc semblé raisonnable de considérer la moyenne arithmétique des deux exposants G et D afin d'obtenir un unique exposant monaural pour chaque auditeur (de la même manière par exemple, que les deux exposants ont été moyennés dans Marozeau et Florentine (2009) pour se ramener à une unique fonction de sonie monaurale).

Les exposants monauraux moyens finalement retenus pour chaque auditeur à 1 kHz et 5 kHz sont présentés dans le tableau 3.2.

Les médianes des exposants individuels à 1 kHz et 5 kHz ont pour valeurs 0.64, soit une valeur très proche de 0.67, la valeur de l'exposant finalement suggérée par Stevens (1972) pour un son pur à 3 kHz. En revanche, les exposants obtenus sont grands comparés aux valeurs moyennes obtenues par

Auditeurs	1 kHz	5 kHz
EP	1.24	1.01
RS	0.65	0.71
FA	0.64	0.56
MV	0.42	0.41
Médiane	0.64	0.64

Tableau 3.2: Exposants monauraux individuels pour les fréquences centrales 1 kHz et 5 kHz et exposants médians sur l'ensemble des auditeurs.

la même méthode dans la littérature. Par exemple, les valeurs de l'exposant moyen α rapportées par Hellman (1981), Hellman et Meiselman (1988) ou encore Collins et Gescheider (1989) par des procédures similaires, étaient respectivement 0.51, 0.46 et 0.58 (pour un son pur à 1 kHz). Une partie de ces différences pourrait être expliquée par le fait que nous utilisons des bruits à bande étroite et non des sons purs. D'autre part, la prise en compte de l'effet d'assimilation suivant la méthode de Cross (1973) a pour effet d'augmenter la valeur de l'exposant de manière plus ou moins importante selon les auditeurs. Cela pourrait donc également participer au fait que nous avons obtenu des exposants plus grands que dans les autres études (ne prenant pas en compte ce biais séquentiel).

3.2 Mesure des fonctions de sonie par production de rapport (*RP*)

Un biais connu, communément appelé "effet de régression" (Stevens et Greenbaum, 1966), désigne le phénomène de compression de l'échelle de la variable de réponse par l'auditeur. Pour l'estimation de grandeur, cela se traduit par une compression de l'échelle de nombres utilisée et donc en une sous-estimation systématique de l'exposant. Il a été suggéré par Cross (1973) qu'une part de l'effet de régression pouvait être attribuée aux effets séquentiels et notamment à l'effet d'assimilation, que nous avons pris en compte précédemment dans l'estimation des exposants pour l'estimation de grandeur. Cependant, l'amplitude de l'effet d'assimilation est trop faible pour permettre d'expliquer la totalité de l'effet de régression rapporté dans la littérature (DeCarlo, 2003). Afin de contrebalancer cet effet, les exposants obtenus par estimation de grandeur peuvent en revanche être combinés avec les résultats obtenus par une méthode de production de grandeur pour laquelle l'effet serait inversé (DeCarlo, 2005). En effet, pour la production de grandeur, le biais se traduirait par une compression des niveaux produits par l'auditeur et impliquerait donc à l'opposé de l'estimation de grandeur, une surestimation (équivalente) systématique de l'exposant. Une méthode un peu différente a cependant été préférée à la production de grandeur pour différentes raisons, dont les principales étaient la rapidité, la non nécessité de connaître les résultats de l'estimation de grandeur au préalable et l'originalité de la méthode. Ainsi, les fonctions de sonie monaurales ont été à nouveau mesurées pour les quatre mêmes auditeurs à l'aide d'une méthode de production de rapport (e.g. Geiger et Firestone (1933), Ellermeier et Faulhammer (2000), Zimmer (2005)).

3.2.1 Méthode

3.2.1.1 Stimuli

Les stimuli étaient strictement identiques à ceux utilisés pour l'estimation de grandeur (bruits filtrés par bandes de tiers d'octave centrées à 1 kHz ou 5 kHz, cf. partie 3.1.1).

La procédure utilisée impliquait la présentation de sons par paires (cf. section 3.2.1.3), le premier étant toujours le son de référence de niveau fixe. Deux niveaux différents ont été étudiés pour le son de référence : 50 dB SPL et 60 dB SPL.

3.2.1.2 Auditeurs

Les quatre auditeurs ayant participé à l'estimation de grandeur (cf. partie 3.1.1) ont également participé à la production de rapport.

3.2.1.3 Procédure

Comme pour l'estimation de grandeur, les sons étaient générés à l'aide d'une carte son Echo Gina 24 (16 bits, 44100 Hz) puis restitués au casque (Sennheiser HD 650) dans une cabine audiométrique.

Les sons étaient présentés par paires, toujours constituées d'un son de référence de niveau fixe, suivi d'un son de comparaison strictement identique, mais de niveau variable. La tâche de l'auditeur était cette fois-ci d'ajuster le niveau du son de comparaison pour qu'il soit perçu 2 fois ou 3 fois plus fort que le son de référence. Au début de chaque bloc, la consigne suivante était rappelée aux auditeurs : *"Des paires de sons vont vous être présentées. Pour chaque paire de sons, un facteur multiplicatif vous sera donné (2 fois ou 3 fois). Il vous est demandé d'ajuster le niveau du son 2 afin qu'il vous paraisse 2 fois ou 3 fois plus fort que le son 1 de référence, en accord avec le facteur en question. Les paires de sons sont présentées en boucle et il vous est proposé de modifier le niveau du son 2 entre chaque présentation. Utilisez la touche "-" pour atténuer le niveau du son 2 et la touche "+" pour l'augmenter. Une fois que le niveau vous paraît satisfaisant, vous pourrez valider votre réponse à l'aide de la touche "Entrée". "*

Chaque mesure commençait donc par la présentation du rapport de sonie à produire (2 fois ou 3 fois plus fort). Le son de référence de niveau fixe (50 ou 60 dB SPL) était ensuite joué, suivi du son de comparaison de niveau variable que l'auditeur avait à ajuster. Pour chaque mesure, le niveau de départ du son de comparaison était tiré de manière aléatoire dans l'intervalle compris entre 5 et 15 dB au-dessus du niveau du son de référence. Les deux sons de chaque paire étaient séparés par un intervalle de 500 ms. La paire de son était ensuite répétée en boucle toutes les 4.5 s.

Dans les 2 s qui séparaient la fin du 2^e son et le début de la nouvelle paire, la possibilité était donnée à l'auditeur de modifier le niveau du son de comparaison. Il pouvait ainsi, soit augmenter le niveau du son de comparaison en pressant la touche "+", soit le diminuer en pressant la touche "-", soit ne pas agir. Le pas de variation du niveau du son de comparaison variait avec le nombre d'inversions, fixé initialement à 4 dB jusqu'à la 1^{re} inversion, à 2 dB jusqu'à la 2^e inversion, à 1 dB jusqu'à la 3^e inversion puis à 0.5 dB ensuite. On notera que le niveau du son de comparaison était limité entre le niveau du son de référence (limite basse) et 90 dB SPL (limite haute). La séquence était répétée ainsi jusqu'à ce que l'auditeur valide sa réponse. Il était possible de valider la réponse, uniquement si la paire de sons avait été écoutée au moins une fois sans modification du niveau.

Au total, l'expérience était constituée de 16 conditions expérimentales (2 fréquences centrales [1 kHz ; 5 kHz] x 2 côtés de présentation [G ; D] x 2 niveaux de référence [50 dB SPL ; 60 dB SPL] x 2 rapports de sonie [2x ; 3x]), répétées quatre fois chacune. Ainsi 4 blocs de 16 mesures ont été réalisés par chaque auditeur, chaque bloc contenant une fois chacune des conditions expérimentales mélangées de manière aléatoire. La durée totale du test était d'environ 1 h par auditeur.

3.2.2 Résultats

Pour chaque condition expérimentale, le niveau produit L_P était défini comme la moyenne arithmétique des niveaux relatifs du son de comparaison sur les quatre répétitions de la mesure. A partir de chaque niveau produit, il était alors possible d'extraire un coefficient directeur β correspondant en théorie, à l'inverse de l'exposant α des fonctions de sonie (modélisées par l'expression $S = k.P^\alpha$). En effet, si le niveau du stimulus produit est exprimé en fonction du logarithme de la sensation, on obtient l'équation suivante :

$$\frac{L_P}{20} = \beta \cdot \log(S) + C, \text{ avec } \beta = \frac{1}{\alpha} \text{ et } C = -\frac{\log(k)}{\alpha} - \log(P_0) \quad (3.2)$$

Un exposant α a ainsi été estimé pour chaque niveau de référence (50 dB et 60 dB) et chaque facteur multiplicatif (2 fois et 3 fois). Soit L_0 le niveau de référence et S_0 la sensation associée. Si multiplier la sensation par F se traduit par une augmentation du niveau de $\Delta L = L_P - L_0$, le niveau produit peut s'exprimer sous la forme :

$$\frac{L_0 + \Delta L}{20} = \beta \cdot \log(F \cdot S_0) + C \quad (3.3)$$

Ainsi,

$$\alpha = \frac{1}{\beta} = \frac{20 \cdot \log(F)}{\Delta L} \quad (3.4)$$

Les valeurs des coefficients α retenues pour chaque niveau de référence (moyenne arithmétique des valeurs pour $F = 2$ et $F = 3$) sont présentées dans le tableau 3.3.

On constate que les pentes obtenues pour le niveau de référence 50 dB SPL sont toujours plus faibles que les pentes obtenues pour le niveau de référence 60 dB SPL. Une ANOVA à mesures répétées sur 3 facteurs (Niveau x Fréquence x Oreille) a été réalisée sur les valeurs de β . Un effet très significatif du niveau de référence [$F(1,3) = 1163.7$; $p < 0.01$] a permis de confirmer l'influence du niveau global sur la valeur de l'exposant. Aucun autre effet ni de la fréquence, ni de l'oreille de présentation n'a été relevé.

Jusqu'à présent, un modèle suivant une simple loi de puissance (Stevens, 1955) a été considéré pour modéliser les fonctions de sonie. Cependant, des résultats plus récents ont montré que la valeur de l'exposant variait avec le niveau, donnant lieu à une nouvelle proposition de modèle : "l'InEx Function" (Buus et Florentine (2002); Florentine et Epstein (2006)). Dans ce modèle, la pente est moins importante pour les niveaux modérés (minimum entre 40 et 60 dB SPL) que pour les niveaux faibles ou forts. En accord avec cette représentation, il serait donc normal que nous obtenions un exposant toujours plus important lorsque la référence est à 60 dB SPL plutôt qu'à 50 dB SPL.

Auditeur	Oreille	1 kHz		5 kHz	
		50 dB SPL	60 dB SPL	50 dB SPL	60 dB SPL
EP	G	0.41	0.50	0.37	0.45
	D	0.39	0.52	0.39	0.48
RS	G	0.38	0.42	0.37	0.53
	D	0.34	0.45	0.37	0.44
FA	G	0.47	0.65	0.58	0.64
	D	0.47	0.70	0.54	0.73
MV	G	0.61	0.85	0.48	0.68
	D	0.51	0.78	0.54	0.66

Tableau 3.3: Coefficients α individuels estimés pour chaque niveau de référence (50 ou 60 dB SPL) aux deux fréquences centrales (1 kHz et 5 kHz)

Cependant, dans la mesure où l'exposant varie avec le niveau global dans les conditions expérimentales de cette expérience, il devait également varier dans le cadre de l'estimation de grandeur. Or, cette variation n'avait pas été prise en compte dans cette première expérience puisque la régression avait été effectuée directement sur l'ensemble des niveaux SPL. Les exposants obtenus pour les deux niveaux de référence ont donc finalement été moyennés afin d'obtenir des résultats comparables pour les deux expériences.

Les exposants α ainsi obtenus pour chaque fréquence centrale et chaque oreille sont représentés dans la figure 3.3.

Des différences dans la valeur de l'exposant sont à nouveau constatées entre les auditeurs. Les auditeurs EP et RS ont obtenu plutôt des valeurs autour de 0.4, tandis que les auditeurs FA et MV ont obtenu plutôt des valeurs autour de 0.6. Sur l'ensemble des conditions, la valeur de l'exposant varie entre 0.39 et 0.71. L'étendue est donc moins grande que pour l'estimation de grandeur (valeurs comprises entre 0.36 et 1.27). A titre de comparaison, Hellman (1981) a obtenu des exposants individuels (par production de grandeur sur sept auditeurs) compris entre 0.5 et 1.2 à 1 kHz et entre 0.54 et 1 à 3 kHz. De la même manière, des exposants compris entre 0.5 et 0.9 pour 10 auditeurs ont été rapportés par Hellman et Meiselman (1988) pour un son pur à 3 kHz. L'amplitude des différences interindividuelles que nous obtenons pour 4 auditeurs est donc en l'occurrence un peu moins grande (sans doute à cause de la taille de l'échantillon), et les valeurs des exposants semblent légèrement plus basses que celles obtenues par les méthodes de production de grandeur. Concernant ce dernier point, il semblerait néanmoins que les méthodes faisant intervenir des jugements de rapport (ou d'intervalle) de sonie aboutissent généralement à des valeurs plus faibles de l'exposant que les méthodes d'estimation ou de production de grandeur (Marks, 1974). Cette remarque sera abordée plus en détail dans la partie 3.3.

D'autre part, en accord avec les résultats de l'analyse de variance réalisée plus haut, on constate que les exposants sont très proches d'une fréquence à l'autre et qu'ils sont identiques pour les deux

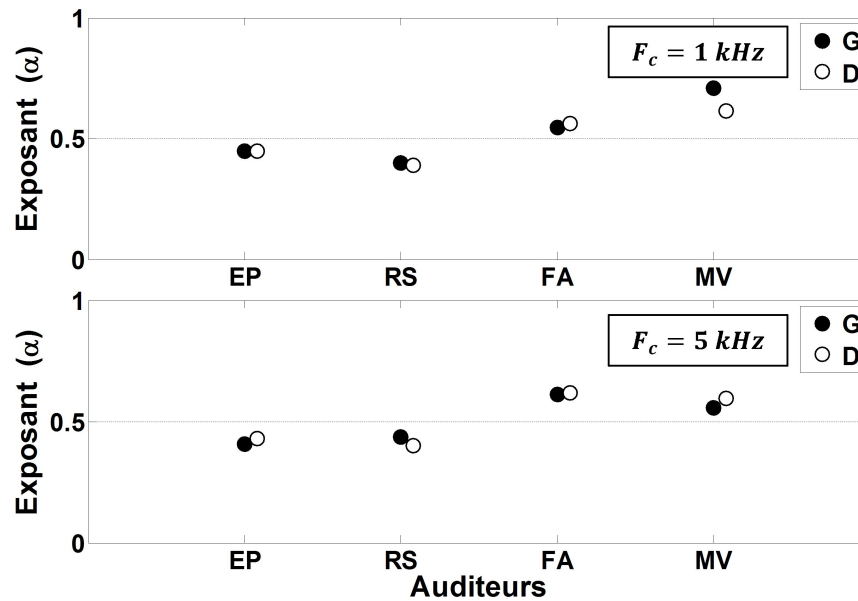


FIGURE 3.3: Exposants des fonctions de sonie obtenus par production de rapport pour chacun des 4 auditeurs. Les symboles pleins représentent les résultats pour l'oreille G et les symboles vides les résultats pour l'oreille D. Les graphiques haut et bas correspondent respectivement aux fréquences centrales 1 kHz et 5 kHz.

oreilles de présentation (sauf peut-être pour MV à 1 kHz). Ces observations (contraires à celles qui avaient été faites pour l'estimation de grandeur) sont davantage en accord avec les résultats de la littérature. Le moyennage des exposants entre les deux oreilles pour aboutir à une unique fonction de sonie monaurale prend alors tout son sens. Les exposants monauraux moyens ainsi obtenus sont présentés dans le tableau 3.4.

Auditeurs	1 kHz	5 kHz
EP	0.45	0.42
RS	0.39	0.42
FA	0.55	0.61
MV	0.66	0.58
Moy.	0.50	0.50

Tableau 3.4: Exposants α des fonctions de sonie monaurales individuelles obtenues à 1 kHz et 5 kHz par production de rapport.

A 1 kHz comme à 5 kHz, les exposants médians sur l'ensemble des auditeurs ont pour valeur 0.50. C'est une valeur inférieure aux exposants moyens obtenus par Geiger et Firestone (1933) (

$\alpha = 0.85$) ou Stevens (1957) ($\alpha = 0.60$) à l'aide d'une méthode de production de rapport, pour des sons purs à 1 kHz. En revanche, des valeurs similaires ont été rapportées par Robinson (1957) ($\alpha = 0.50$) pour un sinus à 1 kHz ou par Churcher *et al.* (1934) ($\alpha = 0.53$) pour une fréquence de 800 Hz, à l'aide de la même méthode. Les exposants estimés à l'issue de cette expérience se situent donc finalement dans une plage de valeurs cohérente au regard des différentes études antérieures.

3.3 Discussion sur les exposants, comparaison des deux méthodes

Les exposants obtenus à l'issue des deux expériences (*ME* et *RP*), à 1 kHz et 5 kHz, sont réunis sur un même graphique (figure 3.4).

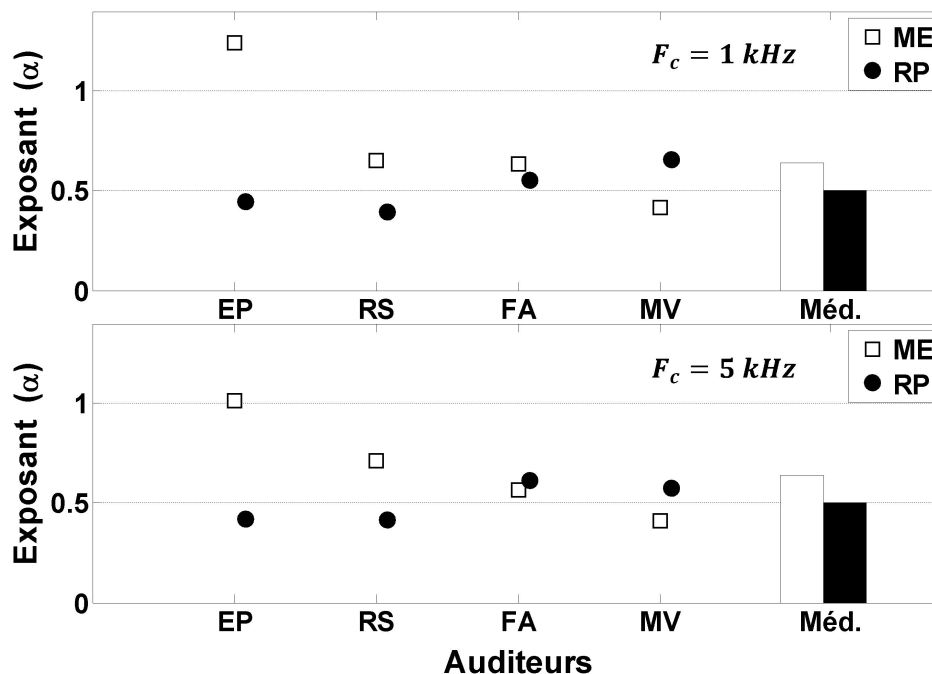


FIGURE 3.4: Comparaison des exposants individuels issus des deux méthodes de mesure *ME* (carrés blancs) et *RP* (cercles noirs). Les barres à droite représentent les exposants médians sur l'ensemble des auditeurs. Les graphiques du haut et du bas distinguent les résultats pour les fréquences centrales 1 kHz et 5 kHz.

Les deux méthodes ont fourni des résultats différents, et surtout ces différences varient fortement d'un auditeur à l'autre. Pour l'auditeur FA, les deux méthodes fournissent des exposants à peu près équivalents. Pour les auditeurs EP et RS, les exposants obtenus par *ME* sont plus grands que ceux obtenus par *RP*, notamment pour l'auditeur EP pour lequel les différences entre les deux méthodes sont particulièrement grandes (facteur 3 entre les deux exposants à 1 kHz). En revanche, pour l'auditeur MV la tendance s'inverse et les résultats obtenus par *RP* sont plus grands que ceux issus de la *ME*. Cela se traduit par des exposants médians (sur l'ensemble des auditeurs) toujours plus grands par la méthode d'estimation de grandeur (0.64 à 1 kHz et 5 kHz) que par la méthode de production

de rapport (0.50 pour les deux fréquences).

La tendance observée est à l'opposé des observations rapportées par les nombreuses études comparant les fonctions de sonie mesurées par estimation et par production de grandeur (e.g. Stevens et Greenbaum (1966); Teghtsoonian et Teghtsoonian (1978)). En effet, tandis que nous obtenons des exposants médians plus grands par estimation de grandeur, "l'effet de régression" prédit une tendance opposée (Stevens et Greenbaum, 1966). D'après l'étude de DeCarlo (2005) (Tab. 1) qui recensait un certain nombre d'expériences ayant consisté à combiner les deux méthodes (*ME* et *MP*), les exposants moyens obtenus par production de grandeur étaient toujours plus grands que les résultats obtenus par estimation de grandeur (quels que soient les stimuli utilisés (bruit, sinus à 1 kHz ou 3 kHz)). DeCarlo (2005), faisait notamment référence aux travaux de Stevens et Guirao (1962), Stevens et Greenbaum (1966), Green *et al.* (1977), Hellman (1981) et DeCarlo (2003). L'effet de régression mis en évidence sur les données moyennes était également constaté de manière systématique pour chaque auditeur (e.g. Hellman (1981); Green *et al.* (1977)). Dans notre cas, les résultats de l'estimation de grandeur sont plus grands pour deux auditeurs sur quatre (EP et RS), ce qui suggère l'implication d'autres facteurs, jouant peut-être un rôle plus important pour ces deux auditeurs.

Un premier élément à prendre en compte est l'augmentation des exposants calculés pour l'estimation de grandeur par la méthode de Cross (1973), qui vise à corriger les effets d'assimilation induits par le contexte local. En effet, ces biais séquentiels allant dans le sens de l'effet de régression, leur prise en compte tend nécessairement à réduire (ou inverser) les différences d'exposant attendues entre les deux méthodes (*ME* et *RP*). De plus, il est intéressant de noter que le recours à la régression multiple selon Cross (1973) impliquait globalement un rehaussement de l'exposant plus important pour l'auditeur EP (augmentation de 0.41 à 1 kHz et de 0.27 à 5 kHz), et dans une moindre mesure pour les auditeurs RS et FA à 1 kHz (augmentation de 0.12), que pour l'auditeur MV. Cela coïncide finalement bien avec l'organisation des différences inexplicables entre les deux méthodes, effectivement plus importantes pour l'auditeur EP. La prise en compte de l'effet d'assimilation suggérée par Cross (1973) fournit donc une première explication.

Une deuxième hypothèse pour expliquer ces résultats est l'obtention d'exposants plus faibles suite à l'utilisation d'une méthode de production de rapport plutôt qu'une méthode de production de grandeur comme dans l'ensemble des études citées. Selon Marks (1974), les exposants obtenus par des méthodes faisant intervenir le jugement ou la comparaison d'intervalles de sonie seraient en effet généralement plus faibles que ceux obtenus par *ME* ou *MP*. La théorie de l'auteur repose sur l'existence de deux échelles de sensations différentes : une échelle qui serait estimée par les méthodes de mesure directe d'une grandeur sensorielle (Type 1, e.g. *ME*, *MP*,...) désignée par l'expression "scale of sensory magnitude" et une échelle qui serait estimée par les méthodes de mesure d'intervalles, de différences, ou de dissimilarités sensoriels (Type 2, e.g. *RE*, *RP*, méthodes catégorielles, bissection,...) désignée par l'expression "scale of dissimilarity". Pour ces deux types de procédures, les résultats de la littérature étaient en accord sur l'expression de la fonction psychophysique sous la forme d'une loi de puissance, mais avec un exposant globalement plus faible pour les procédures de type 2. Dans Marks (1974), la figure 1 rapporte la distributions des exposants des fonctions de sonie issus d'un grand nombre d'études de la littérature. Pour les procédures de type 1, la valeur médiane de l'exposant était 0.54, tandis que pour les procédures de type 2, la médiane valait 0.27. Il y aurait donc un rapport 2 entre les exposants obtenus par les deux types de méthode, et les exposants que

nous avons obtenus par production de rapport seraient donc deux fois moins grands que les exposants que nous aurions obtenus si nous avions opté pour une production de grandeur. Si tel était le cas, tous les auditeurs (sauf EP) obtiendraient alors un exposant équivalent (à l'exposant obtenu par *MP*) supérieur à l'exposant obtenu par *ME* pour les deux fréquences centrales. En considérant que le cas de l'auditeur EP est expliqué par la prise en compte de l'effet d'assimilation (voir ci-dessus), les résultats seraient alors en parfait accord avec la littérature concernant "l'effet de régression".

Ainsi, la combinaison de ces deux hypothèses (réduction de l'exposant pour la *RP* et augmentation de l'exposant pour la *ME* principalement pour les auditeurs EP et RS) est susceptible d'expliquer en grande partie l'inversion des valeurs des exposants par rapport aux résultats de la littérature obtenus par *ME* et *MP*.

L'idée initiale était de moyenniser les résultats des deux expériences suivant la théorie de DeCarlo (2005). En effet, une généralisation du modèle de proportionnalité entre le jugement et la sensation (Stevens (1975)) a été proposée par DeCarlo (2005), suggérant une formulation mathématique du biais (écart à la proportionnalité) impliqué dans les méthodes d'estimation et production de grandeur. Cela a permis entre autre, de fournir une base théorique à la conjecture de Stevens (Stevens, 1975) concernant le moyennage des exposants estimés par chacune des deux méthodes, afin de s'affranchir des biais respectifs (Indow et Stevens, 1966; Hellman et Meiselman, 1988). Le modèle avait également l'avantage de pouvoir fournir une estimation de l'amplitude du biais dans chaque cas étudié. Cependant, étant données les différences entre les exposants issus des deux expériences en contradiction avec l'effet de régression et l'incertitude concernant le rapport des échelles sensorielles (1 :2) suggéré par Marks (1974), le sens attribué au moyennage des exposants par DeCarlo (2005) a été clairement remis en question dans notre cas. Il nous a donc semblé plus pertinent de conserver indépendamment les résultats issus des deux expériences (représentatifs d'une variabilité "réelle" entre les auditeurs, dépendant de la méthode utilisée) pour étudier l'influence de la variabilité des exposants sur les différences interindividuelles dans la DLS.

Finalement, à titre de comparaison, l'ensemble des exposants individuels étaient compris entre 0.42 et 0.66 pour la *RP* et entre 0.41 et 1.24 pour la *ME*. Dans Hellman et Meiselman (1988), des fonctions de sonie individuelles monaurales avaient été mesurées sur un panel de 51 auditeurs, pour des sinus à 1 kHz, en moyennant les résultats obtenus par *ME* et *MP*. Les exposants obtenus variaient entre 0.32 et 1.14. De la même manière, des exposants individuels compris entre 0.4 et 1.1 pour 11 auditeurs en combinant les deux méthodes avaient été rapportés par Stevens et Guirao (1964). L'amplitude de variation des exposants individuels obtenus par *ME* est donc plus proche (que celle des exposants obtenus par *RP*) de l'amplitude rapportée par Hellman et Meiselman (1988) et Stevens et Guirao (1964) en combinant les deux méthodes (exposants "non-biaisés"). D'autre part, les exposants médians sur l'ensemble des auditeurs valaient 0.64 à 1 et 5 kHz pour la *ME* et 0.50 aux deux fréquences pour la *RP*. Ces valeurs sont donc inférieures (*RP*) ou légèrement supérieures (*ME*) à 0.6, valeur standard pour les fonctions de sonie (Stevens, 1955; ISO/R-131, 1959). Elles diffèrent donc de la même manière des valeurs proches de 0.6 couramment trouvées dans la littérature en combinant les deux méthodes. Par exemple Hellman et Meiselman (1988), Stevens et Greenbaum (1966) ou DeCarlo (2003) ont obtenu respectivement des exposants moyens de 0.57, 0.60 et 0.58. L'exposant médian de la *ME* à 1 kHz est en revanche très proche de l'exposant $\alpha = 0.67$ suggéré par Stevens (1972) pour un son pur à 3 kHz. D'une manière générale il semblerait que la prise en compte

de l'effet d'assimilation ait contribué de manière non négligeable à compenser l'effet de régression, rapprochant l'exposant médian obtenu par *ME* des exposants "non-biaisés" rapportés par la littérature. Cependant, ces comparaisons sont données à titre indicatif et ne permettent pas de privilégier une méthode plus qu'une autre, puisque dans tous les cas, les résultats décrits par un échantillon de 4 individus restent peu représentatifs de la moyenne et de la variance réelles de la population.

3.4 Exposants individuels et DLS

L'objectif d'estimer les exposants des fonctions de sonie individuelles était de pouvoir évaluer la contribution de la variabilité des exposants dans la variabilité interindividuelle des DLS.

Dans ce but, les DLS individuelles prédites à partir d'un modèle de sommation de sonie parfaite (modèle utilisé par l'ANSI-S3.4 (2007)) ont été calculées pour les quatre auditeurs EP, RS, FA et MV, permettant de prendre en compte l'effet des exposants de sonie individuels sur la sonie globale. Les DLS prédites ont ensuite été confrontées à la variabilité des DLS mesurées pour ces quatre mêmes auditeurs (cf. chapitre 2). La synthèse binaurale étant générique, les niveaux SPLs aux oreilles considérés étaient les mêmes pour tous les auditeurs. L'unique différence entre les auditeurs dans la prédiction du modèle résidait dans la prise en compte des exposants de sonie individuels. Ainsi, si le rôle de la variabilité des fonctions de sonie était important, la variabilité des DLS prédites (à partir des exposants individuels) devait être proche de la variabilité des DLS mesurées, et inversement.

Le modèle proposé par Sivonen et Ellermeier (2008) permet de représenter précisément les données médianes, mais ne permet pas de rendre compte des différences interaurales de sonie induites par l'exposant de la fonction de sonie. Pour une même différence interaurale de niveau, le modèle suggère de calculer un même niveau monaural équivalent puis de calculer la sonie (en prenant en compte l'exposant). Si la pente est faible, la sonie globale est plus faible et si la pente est forte, la sonie globale est plus forte. Mais comme ici l'objectif était d'égaliser la sonie du son de comparaison avec la sonie du son de référence, l'effet de l'exposant sur la sonie globale est équivalent pour les deux sons et la différence de sonie entre les deux sons revient finalement à la simple différence des niveaux monauraux équivalents. Par cette approche, la différence de sonie entre le son de référence et le son de comparaison est donc indépendante de α . Le modèle rend uniquement compte d'un effet global de la pente sur la perception du niveau du son et non de l'effet "réel" de la pente sur la perception des différences interaurales, ce qui nous intéresse ici pour comprendre les différences interindividuelles dans la DLS.

Par opposition, l'avantage de l'approche de l'ANSI-S3.4 (2007) est que le modèle prend en compte l'effet des sensibilités de chaque oreille dans le calcul de la sonie. Pour une même différence interaurale, si la pente de la fonction de sonie est faible, le rapport de sonie entre l'oreille G et l'oreille D sera faible tandis que si la pente est forte, le rapport de sonie sera grand. En calculant indépendamment une sonie à G et une sonie à D, le modèle rend ainsi compte de l'effet des fonctions de sonie monaurales sur la façon dont les différences interaurales sont perçues. Basé sur la même approche, le modèle de Moore et Glasberg (2007) a été comparé. La modélisation des inhibitions contralatérales avait peu d'effet sur la variabilité de la DLS prédite à 5 kHz lorsque les différences interaurales de niveaux sont grandes : augmentation de la plage de variation des DLS entre les au-

diteurs inférieure à 0.5 dB par rapport au modèle de sommation de sonie parfaite. Nous avons donc choisi de présenter uniquement les prédictions obtenues par le modèle le plus simple.

A titre de rappel, un modèle de sommation de sonie parfaite consiste à calculer une sonie pour l'oreille G et une sonie pour l'oreille D à partir des niveaux SPL aux oreilles (L_G et L_D) puis à sommer les deux sonies pour obtenir la sonie globale N_{Tot} . Soit ΔL le niveau relatif du son de comparaison par rapport au son de référence au PSE. En faisant l'approximation que pour une bande de bruit en tiers d'octave, la fonction de sonie monaurale est une simple loi de puissance de la forme $N = k \cdot P^\alpha$, la sonie globale du son de référence et celle du son de comparaison au PSE peuvent être exprimées de la manière suivante :

$$N_{Tot_{ref}} = k \times \left[10^{\frac{\alpha \cdot L_{G_{ref}}}{20}} + 10^{\frac{\alpha \cdot L_{D_{ref}}}{20}} \right] \quad (3.5)$$

$$N_{Tot_{comp}} = k \times \left[10^{\frac{\alpha \cdot (L_{G_{comp}} + \Delta L)}{20}} + 10^{\frac{\alpha \cdot (L_{D_{comp}} + \Delta L)}{20}} \right] \quad (3.6)$$

Pour chaque auditeur, α était fixé à la valeur de l'exposant mesuré (par *ME* ou *RP*) pour la fréquence centrale considérée (1 kHz ou 5 kHz). La valeur de ΔL qui minimise la différence $\left| \frac{N_{Tot_{comp}}}{N_{Tot_{ref}}} - 1 \right|$ a été ensuite évaluée pour chaque angle d'incidence avec une résolution de 0.1 dB. Cela revenait donc à prédire pour chaque condition expérimentale, le niveau d'égalisation qu'aurait produit l'auditeur (au regard du modèle de sommation de sonie parfaite), si le seul facteur qui le différenciait des autres auditeurs dans le calcul de la sonie globale était son exposant. Les DLS individuelles ainsi prédites par le modèle ($DLS = -\Delta L$) sont présentées dans les figures 3.5 et 3.6 (utilisation des exposants issus respectivement de la *ME* et de la *RP*). Les zones grises représentent la plage de variabilité décrite par les DLS mesurées pour les quatre mêmes auditeurs (cf. chapitre 2).

D'une manière générale, la variabilité induite par les différences d'exposants individuels sur la DLS est faible comparée à la variabilité interindividuelle totale mesurée. Les différences entre les valeurs des exposants individuels sont cependant plus marquées pour les résultats de la *ME* que pour les résultats de la *RP*, ce qui se traduit par un effet plus important des fonctions de sonie sur la variabilité des DLS dans le cas de la *ME* (figure 3.5).

A 1 kHz, les différences interaurales de niveau sont moins grandes donc la variabilité des exposants a une influence plus faible sur les différences interaurales de sonie, et donc sur la variabilité interindividuelle des DLS. Le maximum de différence entre deux auditeurs est observé à 60° d'azimut à 0 et 45° d'élévation (fig. 3.5) où l'écart des DLS prédites entre les auditeurs extrêmes est seulement de 0.7 dB. Pour les exposants issus de la *RP*, l'écart de DLS induit entre deux auditeurs n'excède pas 0.2 dB à cette fréquence. A 5 kHz, les différences interaurales de niveau sont plus marquées, donc les pentes des fonctions de sonie ont un rôle un peu plus important dans la variabilité des DLS. Les différences interindividuelles de DLS induites par les exposants issus de la *ME* sont claires, l'écart entre les deux auditeurs extrêmes atteignant 1.5 à 2.5 dB à 60 et 90° d'azimut. Les exposants issus de la *RP* étant moins dispersés, la variabilité qui en découle reste toujours inférieure à 1 dB entre deux auditeurs, malgré les grandes différences interaurales. On notera que dans le plan sagittal, il n'y a pas ou peu de différences interaurales donc les exposants n'induisent aucune différence de sonie entre les auditeurs. Cependant, la variabilité interindividuelle totale mesurée dans ce plan reste toujours

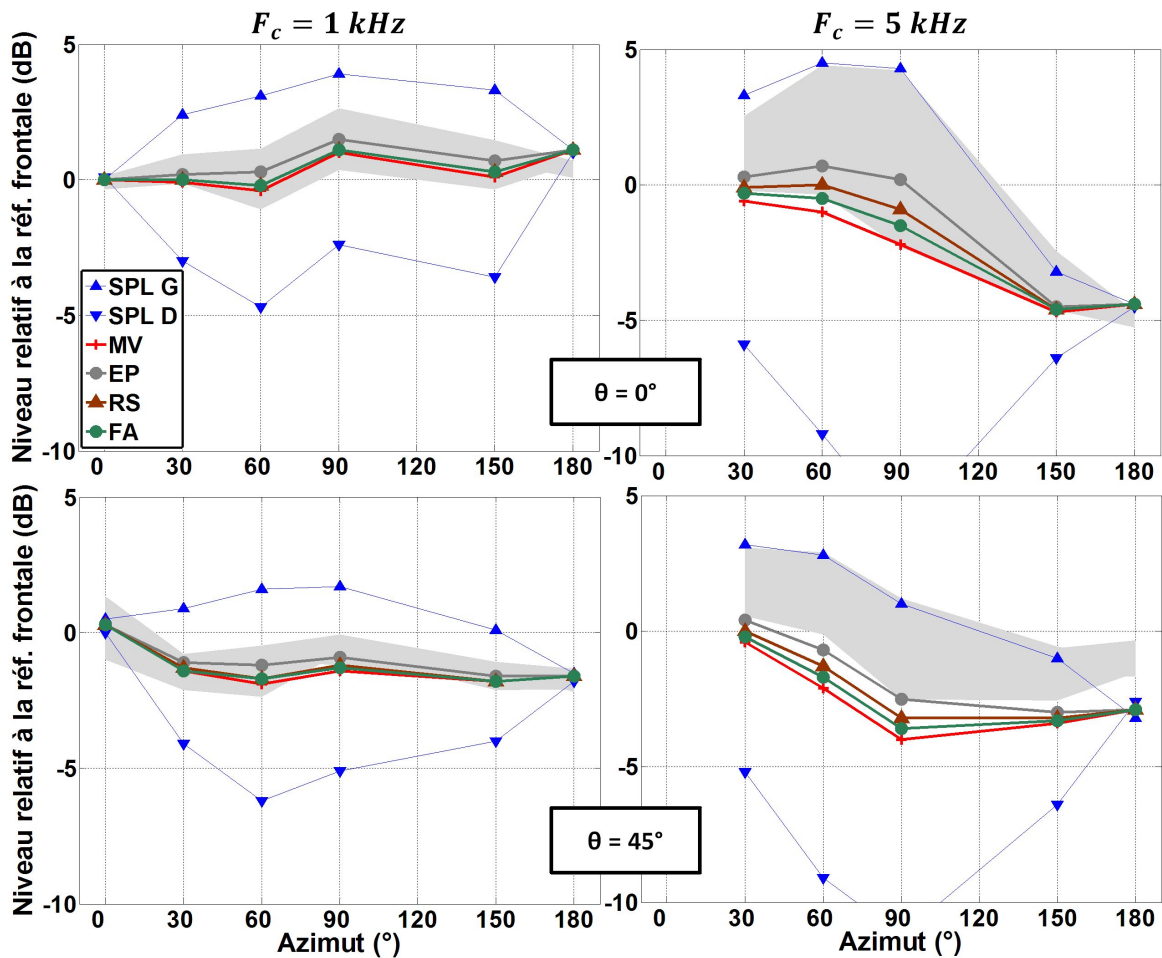


FIGURE 3.5: Prédiction des DLS à partir d'un modèle de sommation de sonie parfaite pour les quatre auditeurs MV, EP, RS et FA, en utilisant les exposants de sonie individuels obtenus par estimation de grandeur (*ME*). La variabilité des DLS mesurées pour ces quatre mêmes auditeurs est représentée par les zones grisées. Les résultats sont présentés à 1 kHz (à G) et à 5 kHz (à D) puis à 0° d'élévation (en haut) et 45° d'élévation (en bas). Les triangles bleus représentent les niveaux SPL aux oreilles.

faible également.

La plage de variation des DLS prédite par les exposants individuels pour les quatre auditeurs EP, RS, FA et MV décrit finalement une proportion comprise entre 0 et 46% par *ME* et entre 0 et 18% par *RP* de la plage de variation totale de la DLS mesurée pour ces mêmes auditeurs. Cependant, en accord avec les résultats issus de la *ME* par exemple, les auditeurs RS et EP ont obtenu respectivement le plus petit et le plus grand exposant. Cela se traduit par des prédictions de la DLS toujours plus petites pour RS et toujours plus grandes pour EP. Or à 5 kHz, lorsque les différences interaurales sont grandes, ces deux auditeurs avaient été à l'inverse choisis pour leurs DLS particulièrement grandes pour RS et faibles pour EP. L'ordre des valeurs de DLS prédites pour les différents auditeurs est donc contradictoire avec l'ordre des valeurs de DLS mesurées pour ces mêmes auditeurs (la même

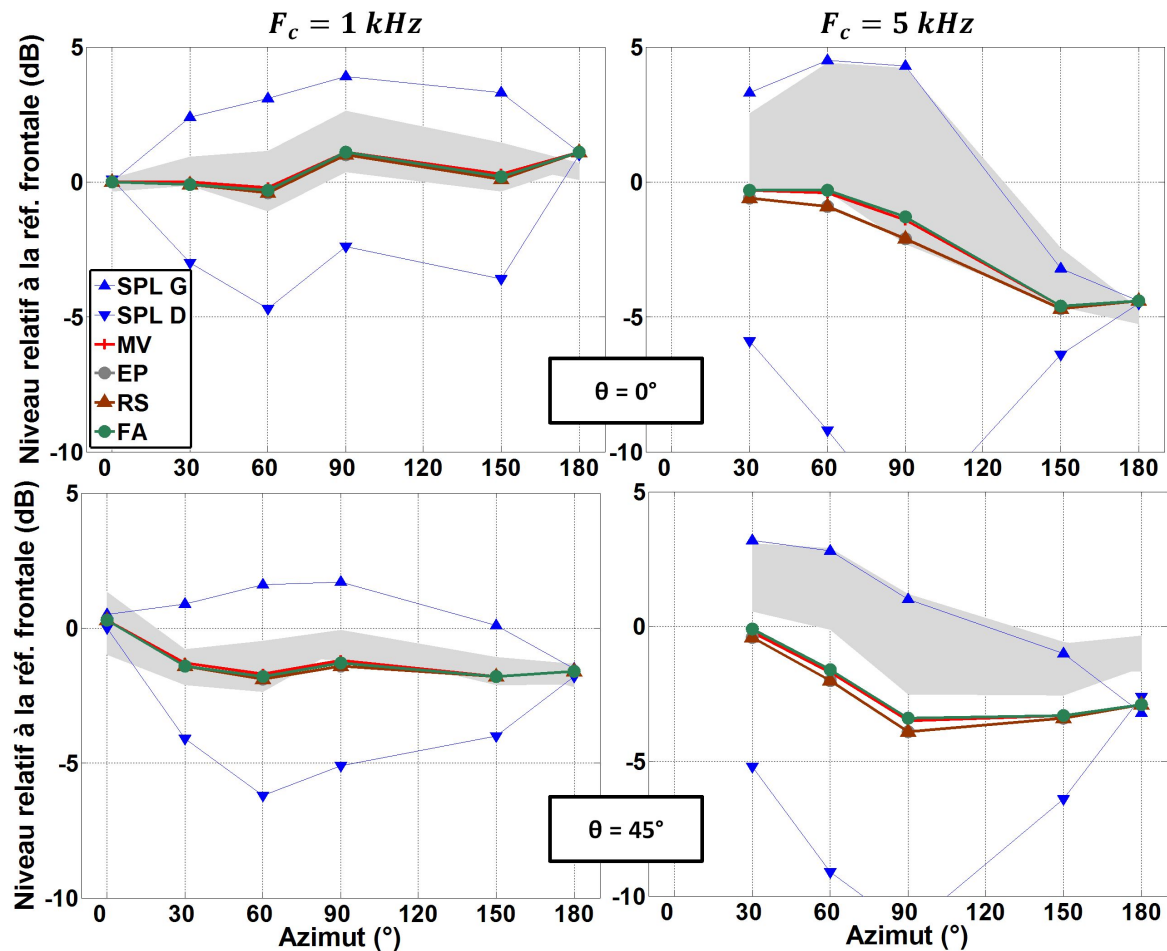


FIGURE 3.6: Prédiction des DLS à partir d'un modèle de sommation de sonie parfaite pour les quatre auditeurs MV, EP, RS et FA, en utilisant les exposants de sonie individuels obtenus par production de rapport (*RP*). La variabilité des DLS mesurées pour ces quatre mêmes auditeurs est représentée par les zones grisées. Les résultats sont présentés à 1 kHz (à G) et à 5 kHz (à D) puis à 0° d'élévation (en haut) et 45° d'élévation (en bas). Les triangles bleus représentent les niveaux SPL aux oreilles.

remarque pouvant être faite à partir des exposants issus de la *RP*). Ainsi, ce n'est pas parce que la plage de variation des DLS prédites décrit 46% de la plage de variation des DLS mesurées que la variabilité des DLS prédites permet d'expliquer 46% de la variabilité des DLS mesurées.

Face à ce constat, deux principales interprétations des résultats sont proposées.

Dans un premier temps, il est possible que l'incertitude sur l'estimation des exposants soit à l'origine d'inversions dans l'ordre des auditeurs, les valeurs d'exposant sont notamment très proches par la méthode de production de rapport ($\alpha = 0.42$ pour EP et RS à 5 kHz). Suivant cette hypothèse, il ne serait alors pas possible de conclure quand à l'implication des exposants de sonie individuels dans la variabilité de la DLS pour les quatre auditeurs considérés. En revanche, en imaginant qu'en réalité l'ordre des exposants respecte bien l'ordre des DLS et joue ainsi un rôle constructif dans la

variabilité de la DLS, il est tout de même possible de quantifier l'influence qu'aurait eue la variabilité des exposants, d'une manière générale. La plage de variation des exposants [0.41 - 1.01] obtenue par *ME* à 5 kHz est représentative des plages de variation "non-biaisées" suggérées par Hellman et Meiselman (1988) ou Stevens et Guirao (1964). Pour une différence de niveau interaurale de 17 dB (ILD maximal à 5 kHz, 90° d'azimut, 0° d'élévation), cette plage de variation des exposants correspond à une plage de variation de 2.4 dB en sonie (pour une sommation de sonie parfaite), soit 42% de la variabilité totale mesurée. Ainsi, même dans ce cas idéal, la variabilité des exposants de sonie individuels permettrait d'expliquer une partie ($< 50\%$) de la variabilité de la DLS, mais pas sa totalité.

Cependant, les exposants obtenus par *ME* et *RP* sont en accord avec les valeurs rapportées dans la littérature pour les mêmes méthodes. En particulier, le fait que les exposants médians obtenus par *ME* (en prenant en compte l'effet d'assimilation) soient proches des valeurs moyennes rapportées par les études combinant *ME* et *MP* et de la valeur $\alpha = 0.67$ suggérée par Stevens (1972) pour un son pur à 3 kHz, ainsi que la cohérence des résultats entre les deux sessions de mesures, confortent dans la confiance à accorder à ces valeurs.

En conséquence, on peut raisonnablement considérer dans un second temps que l'ordre des exposants issus de la *ME* a un sens et qu'il ne respecte pas nécessairement l'ordre des valeurs individuelles de DLS. Ainsi, pour un ILD donné, le rôle de l'exposant des fonctions de sonie individuelles sur le rapport de sonie entre l'oreille G et l'oreille D est vraisemblablement faible (car parfois contradictoire) comparé au rôle des stratégies individuelles adoptée pour combiner les deux sonies (G et D) dans le processus de calcul de la sonie globale. A titre d'exemple, si on se place à nouveau dans la situation où les différences interaurales sont maximales (direction (90°, 0°) à 5 kHz), un ILD de 17 dB était mesuré aux oreilles des auditeurs. Les exposants minimal et maximal obtenus par *ME* valaient respectivement 0.41 (MV) et 1.01 (EP), soit un rapport d'environ 2.5. Les rapports de sonie entre les deux oreilles N_G/N_D induits par les fonctions de sonie individuelles avaient alors pour valeur 2.23 pour MV et 7.22 pour EP. Cette différence n'est donc pas négligeable et se traduit par la prédiction d'une sensibilité en sonie 2.4 dB plus élevée pour EP. En revanche, les mesures ont montré que l'auditeur MV obtenait finalement une DLS 3.7 dB plus élevée que EP, ce qui correspond à une implication contradictoire des exposants individuels dans la variabilité en sonie. Cela montre que les exposants individuels ont une influence notable sur la valeur du rapport de sonie entre les deux oreilles mais que cette influence ne va pas nécessairement dans le même sens que les différences interindividuelles observées en sonie. D'autres facteurs primordiaux (probablement d'ordre cognitifs) sont donc impliqués dans la formation du percept individuel de sonie globale.

3.5 Conclusion

Les fonctions de sonie monaurales ont été mesurées pour quatre des auditeurs ayant déjà participé aux mesures de sonie directionnelles (cf. chapitre 2), à l'aide d'une méthode d'estimation de grandeur puis d'une méthode de production de rapport. Les exposants individuels estimés différaient selon la méthode utilisée, les valeurs et la plage de variation étant globalement plus faibles par *RP* [0.39 - 0.66] que par *ME* [0.41 - 1.24]. Les exposants médians pour chacune des méthodes (0.64 par *ME* et 0.50 par *RP*) étaient en accord avec les résultats de la littérature mais en contradiction avec "l'effet

de régression" normalement observé lorsque les fonctions de sonie sont mesurées à l'aide d'une *ME* et d'une *MP*. Les hypothèses avancées pour expliquer ces différences n'étant pas suffisantes pour privilégier les résultats d'une méthode plutôt qu'une autre ou pour les combiner ensemble, les exposants ont été considérés indépendamment pour chaque méthode dans les analyses de variabilité de la DLS.

Afin de mettre en évidence l'influence de la variabilité interindividuelle des exposants sur la sonie globale en fonction des différences interaurales de niveau, les DLS ont été prédites pour les quatre auditeurs à partir d'un modèle simple de sommation de sonie parfaite, en prenant en compte la valeur des exposants de sonie individuels. Concernant les DLS prédites, la variabilité induite par les différences d'exposants pouvait représenter jusqu'à 46% de la variabilité totale mesurée (pour la *ME*). Mais elle ne permettait pas cependant d'expliquer ce pourcentage de variabilité, étant donné que l'ordre des valeurs de DLS prédites pour chaque auditeur était en contradiction avec l'ordre des valeurs de DLS mesurées pour ces mêmes auditeurs.

Deux interprétations différentes des résultats ont alors été proposées : 1) Soit l'incertitude d'estimation des exposants est à l'origine d'erreurs dans l'ordre relatif des exposants individuels entre eux. Dans ce cas, même en imaginant le cas idéal où les exposants décriraient la même plage de valeurs mais en respectant l'ordre des valeurs individuelles de DLS, la variabilité de la DLS prédite par les exposants ne permettrait pas d'expliquer plus de 42% de la variabilité totale mesurée. 2) Soit on considère que l'estimation des exposants par *ME* est suffisamment précise pour pouvoir se référer avec confiance à l'ordre des valeurs individuelles établi. Dans ce cas, lorsque les différences interaurales de niveau sont grandes, l'exposant de sonie individuel a une influence substantielle sur le rapport de sonie entre les deux oreilles, mais cette influence pour chaque auditeur peut avoir une contribution contradictoire dans la distribution finale des valeurs individuelles de sonie globale. Cela signifierait donc que d'autres facteurs primordiaux (probablement d'ordre cognitifs), intervenant dans la stratégie individuelle de sommation des deux sonies (G et D), seraient alors à l'origine de la majeure partie de la variabilité interindividuelle observée sur les courbes de sensibilités directionnelles en sonie.

Deuxième partie

Sonie binaurale dans un champ acoustique stationnaire complexe

Sonie d'un champ acoustique multi-tonal, composé d'une source bitonale ou de deux sources tonales distribuées

4.1 Introduction

L'objectif de ce chapitre est d'apporter de nouveaux éléments dans la prédiction de la sonie pour des champs acoustiques complexes, à travers une étude faisant intervenir une source multi-tonale ou plusieurs sources tonales distribuées. A partir d'une même méthodologie générale, les objectifs et donc les conditions expérimentales, ont été déclinés en trois grandes parties.

Une première partie est consacrée à l'étude de la sonie directionnelle d'une source de bruit à deux composantes tonales (centrées à 1 kHz et 5 kHz), dans le plan horizontal. Dans une première expérience (cf. Chapitre 2), les sensibilités directionnelles en sonie (DLS) avaient été mesurées sur 17 auditeurs pour chacune des deux composantes tonales. L'objectif ici est de tester l'hypothèse selon laquelle la DLS pour une source complexe multi-tonale pourrait être simplement expliquée par les DLS correspondant à chacune de ses composantes tonales. Dans cette expérience, la sensibilité directionnelle en sonie de la source bitonale est estimée pour 14 auditeurs, chaque composante tonale étant au préalable égalisée en sonie (individuellement) par rapport à l'incidence frontale.

La deuxième partie de l'étude porte sur l'influence de la distribution spatiale de deux sources tonales simultanées sur la sonie. Les travaux menés par Song (2007) pour des angles d'incidence de 10° et 30° , ont montré qu'il n'y avait pas d'effet de la distribution des sources sur la sonie. L'expérience présentée ici vise à tester cette hypothèse pour des grands angles d'incidence (60° et 90°) faisant intervenir des grandes différences interaurales, et donc à valider l'hypothèse dans des situations d'écoute fortement dichotiques. Comme dans la première partie, chaque composante fréquentielle a été au préalable égalisée en sonie pour chaque auditeur, afin de s'affranchir de l'effet de l'angle d'incidence sur la sonie.

Dans les deux premières parties, la référence utilisée est toujours une source bitonale d'incidence frontale. La quantité sensorielle estimée correspond donc à la sonie "relative" par rapport à la sonie de cette référence. Dans la troisième partie, la sonie de la source bitonale d'incidence frontale est directement comparée à la sonie d'une simple source tonale, afin d'obtenir une référence "absolue" de la sonie pour une source bitonale. Cette référence permettra notamment de tester la validité des différents modèles pour prédire la sonie globale d'un champ acoustique multi-tonal, et vérifier si la surestimation des modèles mise en avant par Song (2007) pour un son à deux composantes tonales est à nouveau observée ici.

4.2 Méthode générale

4.2.1 Auditeurs

Quatorze auditeurs (sept hommes et sept femmes âgés de 25 à 50 ans) ayant déjà participé aux précédentes mesures de sonie directionnelle (chapitre 2) ont également participé à cette expérience. Six des auditeurs pouvaient être considérés comme normo-entendants (seuils < 20 dB HL). Les huit autres auditeurs présentaient des pertes auditives légères unilatérales ou bilatérales (seuils ≤ 25 dB HL) à 1 kHz, 2 kHz, 4 kHz ou 6 kHz (fréquences proches de celles des stimuli utilisés dans l'étude). En particulier, l'auditeur RS avait un seuil de 30 dB HL à l'oreille gauche à 1 kHz, l'auditeur CB un seuil de 35 dB HL à l'oreille droite à 2 kHz et l'auditeur SB un seuil de 40 dB HL à l'oreille droite à 6 kHz. L'ensemble des audiogrammes est présenté dans l'annexe A. Les auditeurs étaient indemnisés à hauteur de 10 euros par heure pour leur participation.

4.2.2 Dispositif expérimental

Le dispositif expérimental était strictement identique à celui déjà utilisé pour les mesures de sonie directionnelle (partie 2.1.4). A savoir, la procédure était implémentée dans Matlab ; les sons étaient générés à l'aide d'une carte son Echo Gina 24 (16 bits, 44100 Hz) puis restitués à l'auditeur à l'aide d'un casque Sennheiser HD 650, dans une cabine audiométrique. Un écran était utilisé pour l'affichage, un clavier numérique était utilisé comme terminal de réponse.

4.2.3 Stimuli

Synthèse

Les stimuli étaient des bruits roses filtrés par bandes de tiers d'octave centrées à 1 kHz et 5 kHz (en accord avec les spécifications du standard ANSI-S1.11 (1986)). La durée des stimuli était de 1 s, incluant une croissance et une décroissance en cosinus de 20 ms au début et à la fin du son. Les bruits à bande étroite étaient ensuite filtrés pour simuler des sources réelles au casque, situées à 2 m du centre de la tête de l'auditeur suivant divers angles d'incidence (le principe de la synthèse binaurale générique est décrit dans l'annexe B).

Tous les stimuli utilisés dans cette expérience étaient finalement des sons dichotiques, résultant de la sommation d'une ou de deux composantes fréquentielles centrées à 1 kHz ou 5 kHz, indépendamment spatialisées dans le plan horizontal. Ces stimuli permettaient ainsi de simuler : soit une source bitonale (stimuli A à D), soit deux sources tonales simultanées distribuées dans le plan horizontal (stimuli E à H), soit une simple source tonale (stimuli I et J). La liste complète des différents stimuli synthétisés est détaillée dans le tableau 4.1, en fonction de l'angle d'azimut et du côté de présentation (G ou D) de chaque composante tonale.

Etalonnage : isosonie des composantes

Les différents stimuli ont été étalonnés dans l'optique d'établir l'isosonie à 60 phons de chaque composante dans toutes les configurations de l'expérience (isosonie entre les deux composantes fréquentielles d'une part, isosonie de chaque composante en fonction de l'angle d'azimut d'autre part).

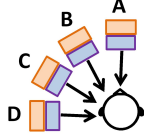
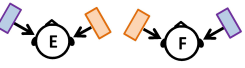
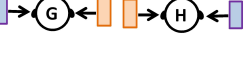
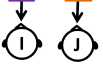
Stim.	1 ^{re} Composante			2 ^e Composante			Légende :
	Fréq.	Az.	Côté	Fréq.	Az.	Côté	
A	1 kHz	0°	-	5 kHz	0°	-	
B	1 kHz	30°	G	5 kHz	30°	G	
C	1 kHz	60°	G	5 kHz	60°	G	
D	1 kHz	90°	G	5 kHz	90°	G	
E	1 kHz	60°	G	5 kHz	60°	D	
F	1 kHz	60°	D	5 kHz	60°	G	
G	1 kHz	90°	G	5 kHz	90°	D	
H	1 kHz	90°	D	5 kHz	90°	G	
I	1 kHz	0°	-	-	-	-	
J	-	-	-	5 kHz	0°	-	

Tableau 4.1: Liste et composition des 10 stimuli synthétisés. La disposition des sources tonales simulées est schématisée dans la colonne de droite pour chaque stimulus.

Pour le stimulus A d'incidence frontale, chaque composante a été calibrée à 60 phones en "équivalent champ diffus". Pour la composante à 1 kHz, cela correspondait donc à une calibration à 60 dB SPL en "équivalent champ diffus" (cf. Annexe C). Pour la composante à 5 kHz, un gain supplémentaire de 0.2 dB était appliqué afin d'obtenir l'isonie théorique entre les deux composantes, en accord avec la norme ISO-226 (2003) et la correspondance entre champ direct et champ diffus mesurée sur la TA Cortex Mk1 (cf. figure 1.12). La calibration des stimuli I et J était strictement identique à celle des composantes fréquentielles de la source bitonale A (respectivement centrées à 1 kHz et 5 kHz).

Pour les autres stimuli (B à H), d'angles d'incidence non nuls, une correction supplémentaire du niveau a été effectuée pour chaque composante, afin d'égaliser la sonie de la composante d'incidence non nulle par rapport à la sonie de la même composante d'incidence frontale à 60 phones. Cela a ainsi permis de compenser au préalable l'effet de l'angle d'azimut sur la sonie pour chaque composante tonale, indépendamment. Les gains utilisés pour cette correction correspondaient aux niveaux d'égalisation obtenus au cours de la première campagne de mesures de sonie directionnelle (cf. Chapitre 2). La figure 4.1 rappelle les DLS individuelles correspondantes pour chaque composante (indépendamment), à 30, 60 et 90°, pour les 14 auditeurs concernés.

Les DLS obtenues pouvaient varier respectivement jusqu'à 3.15 dB et 6.5 dB entre deux auditeurs à 1 kHz et à 5 kHz. Ainsi, étant donnée l'importance des différences interindividuelles qui opéraient (même ordre de grandeur que dans les travaux de Sivonen et Ellermeier (2008)), les corrections appliquées pour chaque composante et chaque azimut ont également été individualisées pour chaque auditeur.

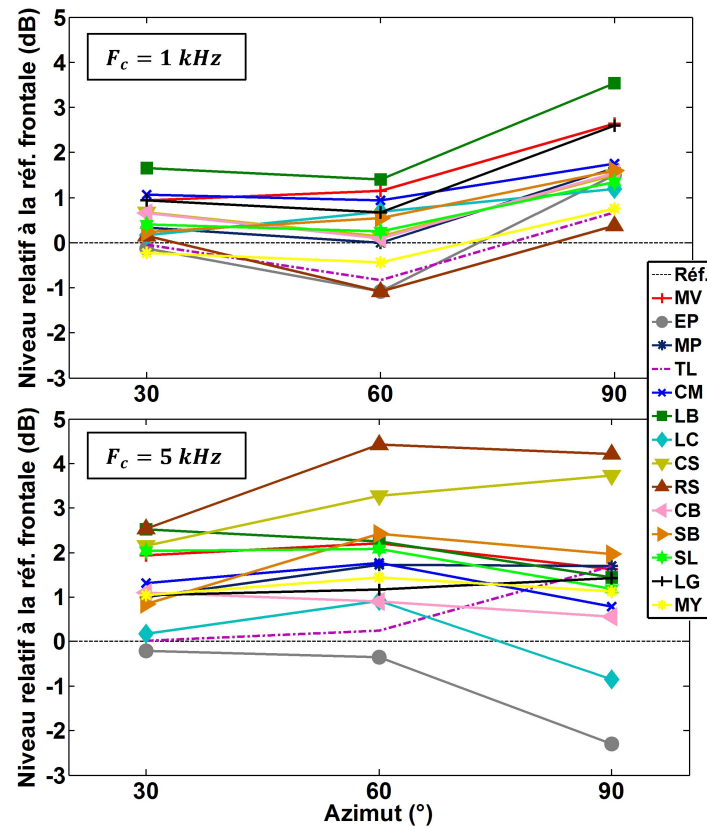


FIGURE 4.1: Sensibilités directionnelles en sonie (DLS) à 30, 60 et 90° d'azimut, mesurées pour chacun des 14 auditeurs, dans le cas d'une source de bruit à bande étroite centrée à 1 kHz (en haut) ou 5 kHz (en bas).

4.2.4 Procédure

L'objectif général de l'expérience était toujours de mesurer le point d'égalité subjective (PSE) en sonie entre un stimulus à une (I et J) ou deux (A à H) composantes tonales d'incidences variables et un stimulus complexe bitonal de référence, d'incidence frontale (A).

Les PSEs ont été mesurés à l'aide d'une procédure adaptative à choix forcé sur deux intervalles (2I-2AFC), suivant une règle 1-down/1-up, permettant d'estimer le point à 50% de réponses positives sur la fonction psychométrique.

Chaque essai consistait à présenter successivement un son de comparaison (de niveau variable) et le son frontal de référence (de niveau fixe). L'intervalle inter-stimuli (ISI) était fixé à 500 ms. L'ordre de présentation des deux sons était choisi de manière aléatoire à chaque essai afin d'éviter les erreurs systématiques de type "Time-order errors" (Hellström, 1979). Il était alors demandé à l'auditeur de désigner le son qu'il percevait comme étant le plus fort (1 ou 2). Selon la réponse de l'auditeur, le niveau du son de comparaison était alors augmenté ou diminué d'un pas de variation de 4 dB jusqu'à la 2^e inversion, de 2 dB jusqu'à la 4^e inversion puis de 1 dB jusqu'à la fin de la mesure (10 inversions). Le niveau relatif du son de comparaison par rapport au son de référence était limité dans l'intervalle [-15 dB, +15 dB]. Le niveau d'égali- sation final pour chaque mesure correspondait à

la moyenne arithmétique du niveau relatif du son de comparaison (par rapport au son de référence) sur les six dernières inversions.

Il y avait en tout 10 conditions correspondant aux différents stimuli à tester (A à I). Pour chaque condition, deux mesures étaient effectuées avec des niveaux relatifs de départ différents pour le son de comparaison (+10 dB ou -10 dB). Chaque mesure était ensuite répétée quatre fois. L'estimation du PSE pour chaque condition, était finalement définie comme la moyenne des 8 niveaux d'égales obtenus (2 niveaux de départ \times 4 répétitions).

Un total de 80 égalisations a été ainsi réalisé par chaque auditeur. Les mesures étaient réparties en 10 blocs de 8 mesures, de manière pseudo-aléatoire (en limitant à 2 le nombre de répétitions d'une même condition au sein du même bloc). Les 8 mesures de chaque bloc étaient entrelacées de manière aléatoire.

L'ordre de présentation des 10 blocs était également tiré au sort pour chacun des auditeurs. Le test était ensuite réparti en 2 séances de 1 h et 1 séance de 30 min.

4.3 Sonie directionnelle d'une source bitonale

La première partie de l'expérience concerne les stimuli A à D, simulant une source bitonale suivant différents angles d'azimuts (0° , 30° , 60° et 90°). Les mesures consistaient à évaluer la sonie directionnelle d'une source bitonale dans le plan horizontal, en s'étant au préalable affranchi de la sonie directionnelle propre à chacune de ses composantes tonales.

4.3.1 Résultats

Les sensibilités directionnelles en sonie individuelles et médianes sur l'ensemble des 14 auditeurs sont représentées dans la figure 4.2, en fonction de l'angle d'azimut. Rappelons qu'une valeur négative indique que le son de comparaison a été perçu moins fort que le son frontal de référence (et inversement).

Pour la condition test (à 0° d'azimut), tous les auditeurs ont été en mesure d'égaliser correctement le stimulus A de référence avec lui même ($|DLS| < 0.7 \text{ dB}$). On constate d'une manière générale que la majorité des auditeurs a obtenu des valeurs proches de zéro à 0° et 90° d'azimut mais que dans la plupart des cas les sources situées à 30° et 60° d'azimut ont été perçues plus fort que la source d'incidence frontale. D'autre part toutes les DLS individuelles sont comprises entre -0.9 et 1.8 dB, la totalité des résultats pouvant être représentée dans un intervalle strictement inférieur à 3 dB. Cela illustre les faibles différences perceptives entre les différents auditeurs et pour les différents angles d'incidence.

Concernant les données médianes sur l'ensemble des auditeurs, les DLS estimées à 0° et 90° d'azimut (respectivement 0.1 et 0.2 dB) sont également très proches de zéro. Un *test des signes* bilatéral sur chaque médiane a permis de montrer que les valeurs obtenues à 0° et 90° peuvent être considérées nulles, contrairement aux deux autres valeurs à 30° et 60° ($p < 0.01$). La médiane augmente donc à 30° et 60° , atteignant la valeur maximale de 1.1 dB à 60° . Cela traduit finalement un effet faible mais significatif de l'angle d'azimut sur la sonie de la source bitonale.

D'autre part, les écarts-interquartiles inférieurs ou de l'ordre du décibel (à 90° d'azimut), traduisent la faible variabilité interindividuelle. Le plus grand écart de niveau entre deux auditeurs est

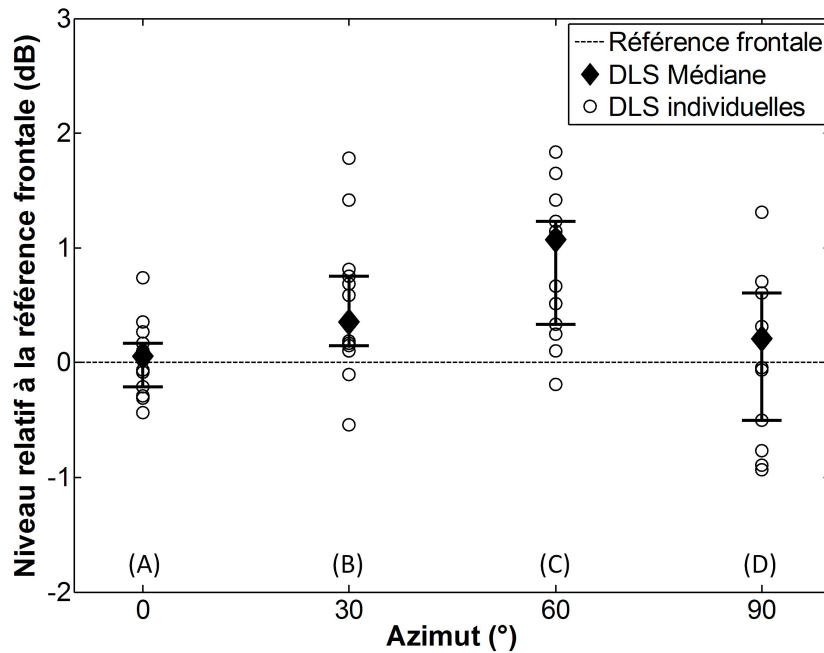


FIGURE 4.2: DLS individuelles (cercles) et médianes (losanges) sur les 14 auditeurs, en fonction de l'angle d'azimut. Les barres d'erreur représentent les écarts interquartiles.

de 2.3 dB (à 30° d'azimut). Cependant, les écarts entre les niveaux d'égali- sation estimés pour les au- diteurs extrêmes (ayant obtenu les DLS min. et max. pour une condition donnée) restent significatifs à 30°, 60° et 90° ($K - S$; $p < 0.01$).

4.3.2 Discussion

Bien que les médianes obtenues sur l'ensemble des auditeurs soient toutes proches de zéro, il existe tout de même un effet faible mais très significatif de l'angle d'azimut sur la sensibilité direc- tionnelle de la source bitonale. Les DLS médianes obtenues à 30° et 60° d'azimut sont en effet plus grandes (atteignant la valeur maximale de 1.1 dB à 60° d'azimut) et significativement différentes de zéro. Étant donné que chaque composante tonale (1 kHz et 5 kHz) a été préalablement égalisée en sonie par rapport à l'incidence frontale pour chaque auditeur (cf. §4.2.3), l'effet de l'azimut mesuré ici est spécifique à la sonie de la source bitonale. Ce résultat réfute donc en partie l'hypothèse de départ, menant à la conclusion que la DLS d'une source bitonale peut-être en grande partie (mais pas entièrement) expliquée par les DLS de chacune de ses composantes tonales.

D'autre part, la variabilité interindividuelle est faible. En effet, les différences interindividuelles observées pour une source à deux composantes (Fig. 4.2) sont fortement réduites par rapport à celles préalablement observées pour chacune des composantes (Fig. 4.1), notamment à 5 kHz. D'après la figure 4.1, à 1 kHz les DLS des différents auditeurs s'étaient respectivement sur un intervalle de 1.88, 2.48 et 3.15 dB à 30, 60 et 90° d'azimut et sur un intervalle de 2.75, 4.77 et 6.50 dB à 5 kHz (mêmes ordres de grandeur que dans l'étude de Sivonen et Ellermeier (2008)). Pour les sources bito- nales, ces mêmes plages de variations deviennent 2.32, 2.02 et 2.25 dB (cf. Fig. 4.2), correspondant à un écart-type d'environ 0.6 dB pour les trois angles d'incidence. A titre de comparaison, pour

chaque condition expérimentale, la variabilité intra-individuelle des niveaux d'égalsation devait être inférieure ou égale à un écart-type de 1.5 dB (critère de validité inter-mesure). La variabilité interindividuelle mesurée pour la source bitonale est donc finalement expliquée en grande partie par l'incertitude de mesure. La variabilité interindividuelle pour les sources à deux composantes est de plus relativement constante avec l'angle d'azimut. Ce n'était pas le cas pour les sources tonales à 1 ou 5 kHz considérées indépendamment, probablement car les différences interindividuelles reflétaient alors des différences de stratégie réelles entre les auditeurs.

Prendre en compte les sensibilités directionnelles individuelles pour chaque composante tonale mène donc à réduire fortement la variabilité interindividuelle dans les mesures de la DLS du son multi-tonal. Cela traduit principalement deux choses. D'une part cela montre à nouveau l'importance du rôle joué par les DLS de chaque composante tonale dans la DLS du son multi-tonal (bien que la totalité de celle-ci ne puisse pas être entièrement expliquée par les DLS de chaque composante). D'autre part cela rend compte de la stabilité et de la robustesse des DLS individuelles mesurées lors de la première expérience, et donc des stratégies individuelles de sommation binaurale, d'autant plus que les deux expériences ont été réalisées à plus d'un an d'intervalle.

4.4 Sonie de deux sources tonales simultanées, distribuées dans le plan horizontal

La deuxième partie de l'expérience concerne les stimuli E à H, simulant deux sources tonales simultanées (centrées à 1 kHz et 5 kHz) et distribuées dans le plan horizontal. L'objectif était d'évaluer l'effet de la distribution des sources sur la sonie pour des grands angles d'incidence ($> 30^\circ$). Pour cela, la sonie des deux sources distribuées a été mesurée à 60° et 90° d'azimut (stimuli E, F, G et H) puis comparée à la sonie des deux sources focalisées suivant les mêmes angles d'incidence (stimuli C et D).

4.4.1 Résultats

La figure 4.3 représente les sensibilités en sonie individuelles (cercles) et médianes (losanges) obtenues sur l'ensemble des 14 auditeurs pour chacune des conditions expérimentales. Les barres d'erreur représentent les écarts interquartiles. Les sources étaient distribuées de manière symétrique par rapport au plan sagittal, suivant des angles de 60° (E et F) ou 90° (G et H), la composante à 5 kHz étant présentée soit à droite (E et G), soit à gauche (F et H).

En accord avec les données individuelles, les sensibilités médianes obtenues sont toujours plus grandes lorsque la composante à 5 kHz est présentée du côté gauche (tendance observée pour 10 auditeurs sur 14). De plus, pour un même côté de présentation, les médianes estimées pour des sources à 60° d'azimut sont toujours plus grandes que les médianes correspondantes obtenues à 90° (tendance observée pour 12 auditeurs sur 14). Une ANOVA à mesures répétées a permis de souligner le caractère significatif des effets observés de l'angle d'incidence [$F(1, 13) = 60.88$; $p < 0.001$] et du côté de présentation des composantes [$F(1, 13) = 21.75$; $p < 0.001$], sans interaction entre les deux facteurs.

L'effet de l'angle d'incidence sur la sensibilité en sonie de deux sources distribuées est en accord avec les mesures de sensibilité directionnelle de la source bitonale effectuées dans la première partie

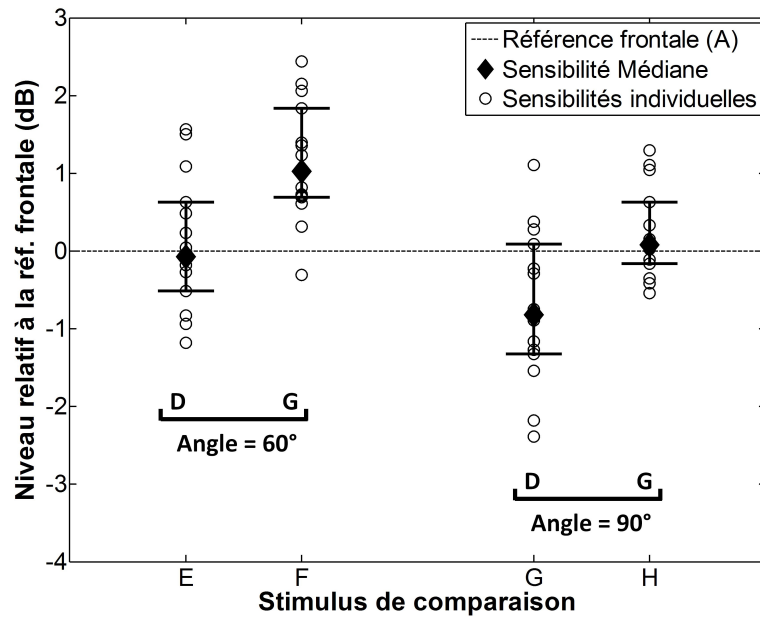


FIGURE 4.3: Sensibilités en sonie individuelles et médianes, en fonction de l'angle d'incidence (60° ou 90°) et du côté de présentation (G ou D) de la composante à 5 kHz.

de l'expérience (§ 4.3). Pour un même côté de présentation de la composante à 5 kHz, la sensibilité lorsque les sources distribuées ont un angle d'incidence de 60° augmente de 0.75 à 0.95 dB par rapport aux sources distribuées à 90° . L'ordre de grandeur est donc le même que dans la première partie, où la sensibilité directionnelle de la source bitonale augmentait de 0.2 dB à 90° d'incidence à 1.1 dB à 60° d'incidence, soit une augmentation de 0.9 dB.

En revanche, la différence systématique des résultats constatée selon le côté de présentation de la source à 5 kHz a suscité une interrogation quant à la pertinence de ce résultat. Ce point est développé dans le paragraphe 4.4.2.

4.4.2 Dissymétrie par rapport au plan sagittal (G/D)

L'obtention d'une sonie plus grande lorsque la source à 5 kHz est présentée à gauche implique une dissymétrie entre la gauche et la droite dépendante de la fréquence, traduisant soit un effet réel du côté de présentation sur la sonie, soit une erreur introduite lors de la restitution des stimuli.

Les signaux restitués à gauche et à droite étaient strictement symétriques entre les conditions E et F (60°) et entre les conditions G et H (90°), seule la correction de la réponse du casque pour chaque oreille restait inchangée quelque soit la condition. Des mesures supplémentaires ont alors été menées pour vérifier la symétrie entre gauche et droite au niveau de la correction de la réponse du casque (cf. Annexe D.1).

La méthode a consisté à convoluer simplement un bruit rose par les réponses impulsionnelles inverses (gauche et droite) du casque couplé avec la tête artificielle. Comme dans le processus de construction des filtres pour la synthèse, un filtrage passe-bande entre 300 Hz et 6 kHz a également été appliqué pour chaque oreille. Les signaux étaient ensuite restitués sur la tête artificielle à l'aide du

casque utilisé dans l'expérience, puis les signaux gauche et droit "aux tympans" étaient enregistrés et analysés en fréquence par bande de tiers d'octave. Vingt répétitions de la mesure ont été effectuées avec un nouveau réglage et un repositionnement complet du casque entre chaque mesure. Le bruit était restitué à 65 dB SPL en équivalent champ diffus. Les résultats obtenus pour les bandes de tiers d'octave centrées à 1 kHz et à 5 kHz sont présentés dans la figure 4.4.

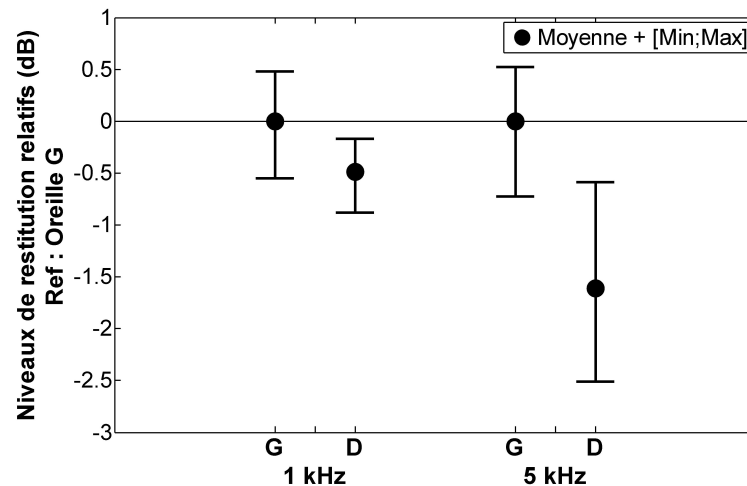


FIGURE 4.4: Niveaux relatifs restitués à G et à D (par rapport aux niveaux restitués à l'oreille G) à 1 kHz et à 5 kHz. Les points et barres d'erreurs représentent la moyenne et les valeurs min. et max. obtenues sur les 20 répétitions de la mesure.

A 1 kHz comme à 5 kHz, les différences obtenues entre les niveaux restitués pour l'oreille gauche et les niveaux restitués pour l'oreille droite sont significatives ($K - S; p < 0.01$). Cependant, tandis qu'à 1 kHz l'écart entre les moyennes reste faible (0.5 dB), à 5 kHz on observe un écart plus conséquent de 1.6 dB. Une analyse sur l'ensemble du spectre révèle en effet une erreur de symétrie systématique dans la correction de la réponse du casque (à partir de 1 kHz environ) qui n'avait pas été décelée plus tôt (cf. Annexe D.1).

Ainsi il est fortement probable que la baisse de 1.6 dB du niveau de restitution pour l'oreille droite à 5 kHz, soit à l'origine de la sous-estimation systématique de la sonie constatée lorsque la composante à 5 kHz était présentée à droite. D'un point de vue perceptif, il n'existe pas à notre connaissance de littérature mettant en avant une telle dissymétrie dans des conditions expérimentales similaires. Dans les travaux de Song (2007) par exemple, aucun effet du côté de présentation n'a été constaté (les composantes tonales étaient centrées à 1 et 3.15 kHz).

Il est important de souligner que cette erreur de niveau était déjà impliquée dans les mesures de sonie directionnelle réalisées au préalable pour chaque composante (Chapitre 2), puisque les réponses du casque utilisées étaient identiques. Cela n'a a priori eu aucune répercussion pour la source à 1 kHz. Néanmoins, il est possible que les résultats à 5 kHz aient été légèrement influencés. En effet, l'atténuation de 1.6 dB du niveau restitué à l'oreille droite, induit une augmentation de 1.6 dB des différences interaurales de niveau lorsque la source est spatialisée du côté gauche de l'auditeur (conditions de l'expérience). Plus la différence interaurale de niveau est grande, plus le niveau de l'oreille contralatérale (niveau le plus faible) semble avoir un poids négligeable dans le calcul de

la sonie globale. Ainsi, l'atténuation du niveau de restitution de l'oreille contralatérale de 1.6 dB a davantage d'influence sur la sonie globale pour les faibles ILDs (source dans le plan sagittal) que pour les grands ILDs (source latérale). D'après le modèle de sommation binaurale proposé par Sivonen et Ellermeier (2008) avec un gain binaural $g = 3 \text{ dB}$, si la source est dans le plan sagittal (ILD nul), une atténuation du niveau de restitution de l'oreille contralatérale de 1.6 dB correspond à une augmentation de la sonie de l'ordre de 0.7 dB. En revanche, dans le cas où l'ILD est maximal (17 dB à 90° d'azimut), cette même atténuation n'a pas de réelle influence sur la sonie ($< 0.05 \text{ dB}$). Pour calculer la DLS, la sonie de la source spatialisée est comparée à la sonie d'une source frontale de référence. Ainsi, dans le plan sagittal, l'atténuation de 1.6 dB a le même impact en sonie pour la source spatialisée et pour la source de référence, il n'y a donc pas d'influence sur la valeur de la DLS. En revanche, plus l'angle d'azimut engendre des différences interaurales importantes et plus l'atténuation a tendance à augmenter la DLS (jusqu'à +0.7 dB à 90° d'azimut dans le plan horizontal, d'après le modèle de Sivonen et Ellermeier (2008)). L'influence de l'erreur dans les réponses du casque sur la DLS était donc d'autant plus importante que les différences interaurales de niveau étaient grandes, mais faible au regard des variations de sonie mesurées.

On notera que cela ne permet pas d'expliquer l'obtention de DLS supérieures aux niveaux SPL à 5 kHz dans le plan sagittal pour les élévations non nulles (résultats non-expliqués, décrits dans la partie 2.2.3). En effet pour les azimuts 0 et 180°, l'erreur dans les réponses du casque ne peut pas a priori, avoir d'effet sur la DLS.

Il est également important de préciser que l'erreur de niveau a pu avoir une légère répercussion sur les valeurs de la DLS mesurée, mais qu'elle a en revanche été prise en compte dans les calculs des modèles (basés directement sur les enregistrements des signaux restitués aux oreilles), et n'a donc en aucun cas pu altérer leurs prédictions.

Finalement, afin de vérifier l'hypothèse selon laquelle l'erreur de niveau dans la correction du casque à 5 kHz est à l'origine de la dissymétrie observée en sonie et conclure quant à l'effet de la distribution des sources sur la sonie, des nouvelles fonctions de transfert du casque (couplé avec la tête artificielle) ont été mesurées et contrôlées dans un test de restitution (cf. Annexe D.2). La partie suivante (§ 4.4.3) présente la répétition de l'expérience sur 6 auditeurs en utilisant la nouvelle égalisation du casque.

4.4.3 Répétition de l'expérience avec une nouvelle égalisation du casque

4.4.3.1 Méthode

Auditeurs Six des auditeurs impliqués dans l'expérience principale (MV, SL, LB, EP, TL et SB) ont également participé à la répétition des deux premières parties de l'expérience (correspondant aux parties 4.3 et 4.4).

Stimuli Seuls les stimuli A à D et E à H étaient concernés dans les deux parties de l'expérience répétées. La synthèse et la calibration étaient strictement identiques à ce qui était décrit dans la méthode générale, à l'exception de l'égalisation de la réponse du casque impliquée dans le processus de synthèse binaural. En effet, l'égalisation consistant à convoluer le signal par la fonction de transfert inverse du casque utilisait ici les nouvelles fonctions de transfert mesurées (cf. Annexe D.2),

respectant la symétrie entre les deux oreilles.

Procédure Huit conditions correspondant aux huit stimuli (A à H) ont été investiguées. Pour chaque condition, deux mesures étaient effectuées avec des niveaux relatifs de départ différents pour le son de comparaison (+10 dB ou -10 dB). Chaque mesure était ensuite répétée deux fois (au lieu de quatre dans la méthode générale). L'estimation du PSE pour chaque condition, était finalement définie comme la moyenne des 4 niveaux d'égalisation obtenus (2 niveaux de départ \times 2 répétitions). Le total de 32 mesures était ensuite réparti en 4 blocs de 8 mesures de manière aléatoire. La durée totale de cette expérience supplémentaire était d'environ 1 h par auditeur. Aucune mesure supplémentaire n'a été nécessaire pour respecter les critères de validité intra- et inter-mesures initialement fixés.

Le reste de la procédure ainsi que le dispositif expérimental étaient strictement identiques à ce qui a été décrit dans la méthodologie générale (§ 4.2).

4.4.3.2 Résultats

La figure 4.5 rassemble les sensibilités directionnelles en sonie mesurées pour la source bitonale (stimuli A, B, C et D) d'une part et les sensibilités en sonie mesurées pour les deux sources distribuées à 60 ou 90° d'azimut (stimuli E, F, G, et H) d'autre part. Les résultats sont présentés strictement de la même manière que pour l'expérience principale (figures correspondantes : 4.2 et 4.3).

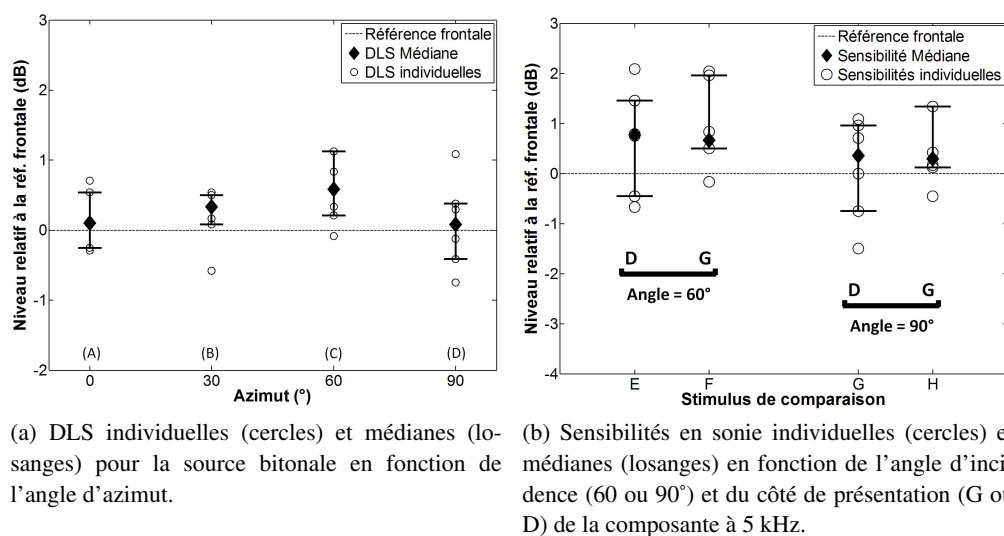


FIGURE 4.5: Sensibilités en sonie (par rapport à la source bitonale A) mesurées pour l'ensemble des stimuli A, B, C, D (fig. a) et E, F, G, H (fig. b).

La figure 4.5a illustrant l'effet de l'angle d'azimut sur la sonie de la source bitonale pour les 6 auditeurs, n'est pas sans rappeler l'évolution des DLS médianes qui avaient déjà été mesurées pour 14 auditeurs (représentées dans la figure analogue 4.2). En effet, les DLS médianes à 0° et 90° d'azimut sont à nouveau très proches de zéro (valant 0.1 dB pour les deux angles) et augmentent légèrement à 30 et 60°, atteignant la valeur maximale de 0.6 dB à 60°. En revanche, contrairement à ce qui avait été montré dans la partie principale, les valeurs obtenues à 30 et 60° ne sont pas

statistiquement différentes de zéro [*TDS, N.S.*] et l'effet de l'angle d'azimut sur la sonie n'est donc plus significatif. Il est toutefois possible que la réduction de la saillance de l'effet soit due à la petite taille de l'échantillon d'auditeurs dont nous avons disposé.

Concernant les sensibilités en sonie pour les sources distribuées (figure 4.5b), on constate que pour un angle d'incidence donné (60 ou 90°), les valeurs médianes sont très similaires, que la composante à 5 kHz soit présentée à gauche ou à droite. L'égalité des données pour chaque angle d'incidence (entre E et F et entre G et H) a été confirmée par deux tests des signes [*N.S.*]. La symétrie par rapport au plan sagittal est donc désormais bien respectée. Concernant l'effet de l'angle d'incidence sur la sonie, les médianes obtenues à 60° d'azimut sont légèrement plus élevées que celles obtenues à 90°, mais cette différence n'est pas significative [*TDS, N.S.*], en accord avec les résultats obtenus pour la source bitonale (stimuli C et D).

Jusqu'à présent, les sensibilités en sonie ont été calculées en référence à la sonie de la source bitonale frontale. Cela revient donc ici à comparer la sonie des deux sources distribuées (à 60° ou 90°) avec la sonie des deux sources focalisées d'incidence frontale. Cependant, un effet de l'azimut sur la sonie a été mis en évidence dans la partie 4.3, spécifique à la source bitonale. Afin de pouvoir conclure quant à l'effet réel de la distribution des sources sur la sonie à 60° et à 90° d'azimut, il faut donc s'affranchir au préalable de l'effet directionnel mesuré pour la source bitonale. Ainsi, la sonie des deux sources distribuées à 60° (E et F) a été directement comparée à la sonie des deux sources focalisées à 60° (C), de la même manière que la sonie des deux sources distribuées à 90° (G et H) a été comparée à la sonie des deux sources focalisées à 90° (D).

Les différences de sensibilité en sonie réelles entre "sources distribuées" et "sources focalisées" ont été calculées pour chaque auditeur. Cela impliquait le calcul individuel des quatre différences de sensibilité en sonie suivantes : $E - C$, $F - C$, $G - D$ et $H - D$. Les valeurs médianes de ces différences et les écarts interquartiles sont représentés dans la figure 4.6. L'écart des données médianes à la valeur zéro correspond à l'effet de la distribution des sources sur la sonie.

On constate qu'à 60° d'incidence, comme à 90° d'incidence, toutes les valeurs médianes sont très proches de zéro (< 0.4 dB), indépendamment du côté de présentation de la composante à 5 kHz. D'après les tests réalisés pour chaque condition, aucune de ces valeurs n'est significativement différente de zéro [*TDS, N.S.*], ce qui traduit l'absence générale d'effet de la distribution spatiale des sources sur la sonie.

4.4.4 Distribution des sources et ségrégation

D'après les résultats finaux de cette partie de l'expérience, il n'y a pas d'effet de la distribution spatiale de deux sources tonales sur la sonie globale. Une question qui se pose alors, est de savoir quels percepts sous-jacents d'intégration ou de ségrégation des composantes sont impliqués. Plus précisément, est-ce que l'auditeur perçoit les deux composantes fréquentielles toujours groupées ou toujours séparées, indépendamment de leur configuration spatiale (focalisée ou distribuée) ? Ou bien, dans le cas où les composantes seraient perçues groupées lorsqu'elles sont focalisées dans une direction de l'espace, est-ce que leur distribution spatiale provoque leur ségrégation ?

La source bitonale frontale (A) était constituée de deux bruits à bande étroite centrés à 1 kHz et 5 kHz. Les deux composantes pouvaient alors être intégrées pour former un son complexe ou

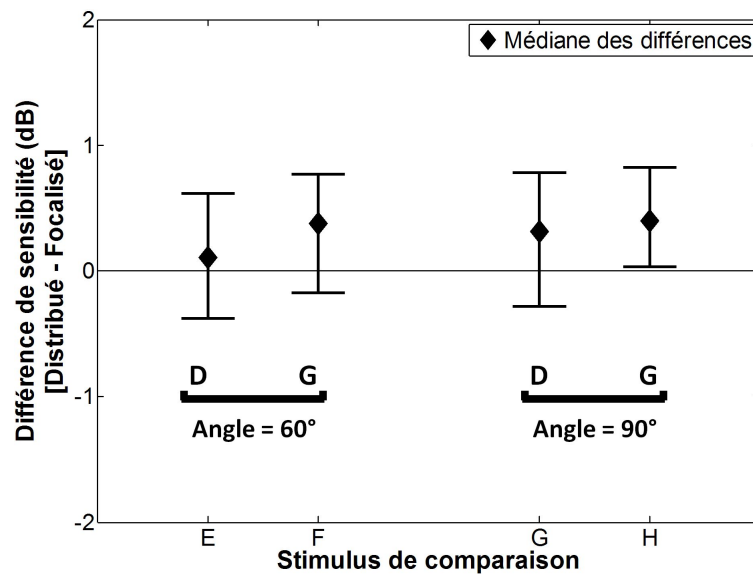


FIGURE 4.6: Différence de sensibilité médiane (Distribué – Focalisé), en fonction de l’angle d’incidence (60 ou 90°) et du côté de présentation (G ou D) de la composante à 5 kHz.

bien ségréguées et perçues comme deux objets distincts. De nombreux facteurs peuvent affecter la fusion des deux composantes comme le contexte tonal et temporel dans lequel elles sont présentées (Bregman (1990)). Dans notre cas, les deux composantes simultanées partageaient la même attaque, la même fin, et aucune des deux composantes n’était présentée seule avant ou après la présentation du son complexe (Bregman et Tougas, 1989). En se basant sur des écoutes critiques, nous pensons que malgré leur écart en fréquence important, les composantes étaient plutôt intégrées par l’ensemble des auditeurs. Concernant les stimuli (E, F, G, H) où chaque composante était spatialisée suivant un angle d’incidence différent, il est alors intéressant de se demander si les indices binauraux (ITD, ILD, contenu spectral) propres à chaque composante étaient susceptibles d’entraîner la ségrégation des deux sources.

David *et al.* (2014) ont montré que des faibles différences spectrales (ordre du dB) correspondant à la réponse de la salle ou au filtrage de la tête de l’auditeur suivant la position de la source, étaient suffisantes pour provoquer de la ségrégation de flux obligatoire pour des signaux de bruit proches de la parole. Cela rejoignait entre autre les résultats de Middlebrooks et Onsan (2012), montrant la possibilité de provoquer une ségrégation volontaire en champ libre, uniquement à partir de différences spectrales subtiles induites par les HRTFs (azimuts inférieurs à 10°). De nombreuses autres études ont également mis en évidence l’importance des indices spatiaux dans le groupement d’éléments au cours du temps, notamment certaines études sur l’intelligibilité de la parole en présence de plusieurs locuteurs (e.g. Darwin et Hukin (1999), Kidd *et al.* (2005)). Les indices binauraux contenus dans les HRTFs constituent donc des indices efficaces pour le groupement séquentiel. Cependant, ces indices semblent avoir peu d’influence sur la ségrégation simultanée.

En effet, plusieurs études ont montré que, pour des signaux stationnaires simultanés ou pour des signaux non-stationnaires mais sur des échelles de temps très courtes, l’ITD avait une influence plus faible que les autres indices de ségrégation tels que l’harmonicité, la continuité spectro-temporelle,

l'attaque ou la modulation commune (e.g. Culling et Summerfield (1995), Hukin et Darwin (1995), Darwin et Hukin (1997), Schwartz *et al.* (2012)). Dans notre cas, les deux composantes diffèrent uniquement par leur hauteur fréquentielle. A titre d'exemple, l'étude de Hukin et Darwin (1995) a mis en évidence que l'attribution d'un ITD différent ($-666 \mu s$) à l'harmonique 500 Hz de la fréquence fondamentale d'un son de voyelle, fréquence proche du premier formant, n'était pas suffisant pour la séparer du reste de la voyelle (auquel un ITD de $+666 \mu s$ était appliqué). Ce résultat traduisant le fait que l'ITD n'est pas (ou peu) pris en compte lors du groupement fréquentiel a également été rapporté par Darwin et Chris (1997) et Culling et Summerfield (1995). Dans Culling et Summerfield (1995), quatre bandes de bruit de 150 Hz (centrées à 225, 625, 975, et 1925 Hz) étaient sommées et correspondaient aux deux premiers formants (F1 et F2) de deux voyelles chuchotées différentes (paire de voyelles différente selon la combinaison des formants). Ainsi, selon la façon dont les formants étaient groupés, l'auditeur pouvait percevoir soit un seul son complexe, soit deux voyelles différentes. L'étude a montré que l'identification de deux voyelles distinctes n'était pas améliorée lorsque les deux composantes d'une même voyelle avaient un ITD commun de $-390 \mu s$ et que les deux autres avaient un ITD commun de $+390 \mu s$, par rapport à la situation où les quatre composantes étaient regroupées avec un ITD commun de $+390 \mu s$. L'identification était en revanche rendue possible lorsque chaque voyelle était exclusivement présentée à une oreille différente. Cela montre donc que le partage d'un ITD commun n'est pas suffisant pour pouvoir séparer un groupe de composantes fréquentielles du son complexe.

Pour les stimuli utilisés dans notre étude (proches de ceux utilisés par Culling et Summerfield (1995)), il est donc probable que la différence d'ITD entre les deux composantes fréquentielles ne soit pas un indice de ségrégation suffisant. En basse fréquence, l'ITD est le principal indice de localisation (Wightman et Kistler, 1992), l'ILD étant très faible (faible effet de masque de la tête pour les grandes longueurs d'onde). En revanche, en hautes fréquences, l'ILD est souvent considéré comme l'indice de localisation principal ("duplex theory"). Considérer uniquement l'effet de l'ITD ne serait donc pas représentatif de l'ensemble des indices spatiaux contenus dans les HRTFs, bien que pour des sons complexes l'ITD de l'enveloppe temporelle joue également un rôle important dans la localisation en hautes fréquences, même pour des ILDs nuls (e.g. Macpherson et Middlebrooks (2002)). Peu d'expériences à notre connaissance ont étudié l'implication de l'ILD comme indice pour la ségrégation simultanée. D'après Beerends et Houtsma (1986), lorsque deux harmoniques consécutifs de deux fondamentales différentes étaient présentés simultanément, les auditeurs n'étaient pas plus performants pour déterminer les fondamentales lorsque chaque note était présentée à une oreille différente que lorsque un harmonique de chaque note était distribué à chaque oreille. De la même manière, l'étude de Darwin et Ciocca (1992) a montré que l'effet d'un harmonique désaccordé sur la hauteur tonale d'un son complexe était le même, que l'harmonique en question soit présenté à la même oreille ou à l'oreille opposée par rapport au reste du son complexe harmonique. Dans ces deux exemples, le fait de séparer physiquement les objets en les présentant à des oreilles différentes est analogue à l'application d'ILDs infinis communs à l'ensemble des éléments d'un même objet. L'ILD semble donc avoir une influence négligeable dans la ségrégation simultanée de différentes composantes fréquentielles pour des sons purs. En revanche, dans Culling et Summerfield (1995), utilisant des bruits à bande étroite (et non de sons purs), l'application d'un ILD commun infini aux deux formants d'une même voyelle (présentation à la même oreille) permettait la ségrégation (contrairement à l'ITD).

D'autre part, une étude menée par Hirvonen (2005) pour des stimuli similaires à ceux utilisés

dans notre expérience, montre que nous sommes en présence d'un écart fréquentiel suffisant pour permettre la ségrégation des deux composantes tonales. L'auteur a utilisé des signaux composés de deux bruits de largeur de bande égale à 1 ERB. Il est apparu que, lorsque le bruit le plus grave était centré à une fréquence inférieure à 1.5 kHz, l'écart minimal conduisant à la ségrégation des deux bandes était de l'ordre de 1 ERB. Hirvonen (2005) met ainsi en évidence la possibilité de ségréguer volontairement les composantes tonales dans les conditions expérimentales de notre étude, mais cela n'implique pas néanmoins la ségrégation "obligatoire" des composantes (Bregman, 1990). On notera que les auditeurs MV et EP avaient une parfaite connaissance de la composition des différents stimuli, et que cela pouvait peut-être les amener à ségréguer plus facilement chacune des deux composantes, lorsque leur attention était portée sur ce point. Cette connaissance supplémentaire ne se traduit pas en tous cas dans les résultats en terme de sonie par rapport aux autres auditeurs.

Finalement, lorsque les deux sources étaient focalisées en incidence frontale, l'auditeur devait percevoir plutôt un seul son (intégration), et lorsque les deux sources étaient distribuées dans le plan horizontal suivant différents angles d'incidence, nous pouvons seulement faire l'hypothèse que les sources étaient plus souvent perçues séparément que dans le premier cas. Cependant, en se basant sur les ITDs, cette différence est vraisemblablement faible (e.g. Culling et Summerfield (1995)). Les ILDs ont peut-être plus d'influence sur la ségrégation des composantes fréquentielles (Culling et Summerfield, 1995), bien que certaines études arrivent à des conclusions contraires pour des sons purs (Beerends et Houtsma (1986), Darwin et Ciocca (1992)). On notera enfin qu'il était demandé à l'auditeur de bien "considérer le son dans son ensemble" pour porter son jugement, ce qui a pu jouer en faveur du percept d'intégration des composantes. Les résultats en sonie suggèrent que les auditeurs percevaient toujours un son unique (intégration), que les composantes tonales à 1 kHz et 5 kHz soient coïncidentes en incidence frontale ou distribuées suivant différents azimuts. Cette suggestion est plutôt cohérente avec les différents constats et résultats de la littérature. Cependant, une expérience supplémentaire aurait été nécessaire pour pouvoir conclure de manière certaine quant au percept prédominant (intégration ou ségrégation) pour chaque configuration des stimuli.

4.4.5 Conclusion

Un protocole expérimental permettant de s'affranchir totalement des effets de l'angle d'incidence sur la sonie et de mettre en évidence uniquement les effets causés par la distribution spatiale des sources a été mis en place.

Les sensibilités en sonie ont été mesurées dans le cas où les deux sources tonales étaient distribuées dans le plan horizontal à $\pm 60^\circ$ (E et F) ou $\pm 90^\circ$ (G et H) d'azimut. Suite au constat d'une dissymétrie en sonie par rapport au plan sagittal (§ 4.4), des mesures supplémentaires ont permis de déceler une différence de niveau entre les deux oreilles à 5 kHz lors de la restitution. L'hypothèse selon laquelle l'erreur de niveau (localisée dans les fonctions de transfert du casque) permettait d'expliquer la dissymétrie en sonie a alors été avancée.

Afin de vérifier cette hypothèse, les deux premières parties de l'expérience ont été répétées sur 6 auditeurs, en utilisant des nouvelles fonctions de transfert du casque (contrôlées au préalable). Les résultats ont montré que lorsque la symétrie était bien respectée lors de la restitution, la dissymétrie en sonie n'apparaissait plus. Cela a donc confirmé l'hypothèse de départ et permis de conclure quant à l'absence d'effet de la distribution spatiale des sources sur la sonie globale pour des grands angles d'incidence (60° et 90°). Cette conclusion constitue une extension des résultats déjà rapportés par

Song (2007) pour des angles de 10° et 30° .

4.5 De la sonie "relative" à la sonie "absolue" d'un champ acoustique multi-tonal

Jusqu'à présent la sonie globale du champ bitonal étudié a été évaluée suivant différentes configurations des deux sources tonales, mais toujours de manière relative à la sonie de la source bitonale de référence frontale (A). Cependant, dans la volonté d'évaluer la capacité des différents modèles à prédire la sonie "absolue" d'un champ acoustique complexe, il est nécessaire de comparer par exemple la sonie de la référence bitonale (A) à la sonie d'une seule composante tonale d'incidence frontale (situation diotique simple, dans le domaine de validité des modèles). Cela permettra également de vérifier l'hypothèse avancée par (Song, 2007), selon laquelle les différents modèles auraient tendance à surestimer la sonie lorsque le champ acoustique est composé de deux composantes fréquentielles.

4.5.1 Comparaison de sonie entre une source bitonale (f_1+f_2) et une source tonale (f_1 ou f_2)

Cette partie concerne les stimuli I et J et vise à comparer la sonie globale de la source bitonale de référence (A) avec la sonie de chacune de ses composantes tonales d'incidence frontale, centrées à 1 kHz (I) ou à 5 kHz (J).

4.5.1.1 Résultats

Les sensibilités en sonie individuelles estimées pour chacune des deux sources tonales I (1 kHz) et J (5 kHz) sont présentées dans la figure 4.7.

La présentation des résultats pour chaque auditeur permet notamment de mettre en avant l'importance de la variabilité interindividuelle. En effet la différence de sensibilité entre deux auditeurs peut atteindre 9.1 dB à 1 kHz (entre les auditeurs CB et RS) et 5.7 dB à 5 kHz (entre les auditeurs LC et SL). De plus, on constate que la variabilité intra-individuelle est également particulièrement élevée. L'intervalle entre les niveaux d'égaleisation minimal et maximal sur les différentes répétitions de la mesure est en moyenne 1.5 à 2 fois supérieur pour les conditions I et J que pour toutes les autres conditions expérimentales. Par ailleurs, six des quatorze auditeurs ont dû participer à au moins une session de mesures supplémentaire pour que les niveaux d'égaleisation entrent dans les critères de validité des mesures que nous avons fixés.

On peut également noter que pour la plupart des auditeurs, la source à 5 kHz a été jugée plus forte que la source à 1 kHz, hormis pour LG, TL, RS, EP et MY qui ont cependant perçu des niveaux relativement proches pour les deux fréquences. Certains auditeurs semblent même avoir perçu la sonie de la source à 5 kHz à des niveaux proches de la sonie de la référence bitonale (voire à un niveau strictement égal pour l'auditeur SL).

La figure 4.8 résume les sensibilités médianes estimées sur l'ensemble des auditeurs pour chaque stimulus de comparaison (I ou J).

Au point d'égalité subjective en sonie, les sources tonales à 1 kHz et 5 kHz ont été ajustées respectivement à 5.6 dB et 2.8 dB au-dessus de leur niveau dans la source bitonale. Cependant les

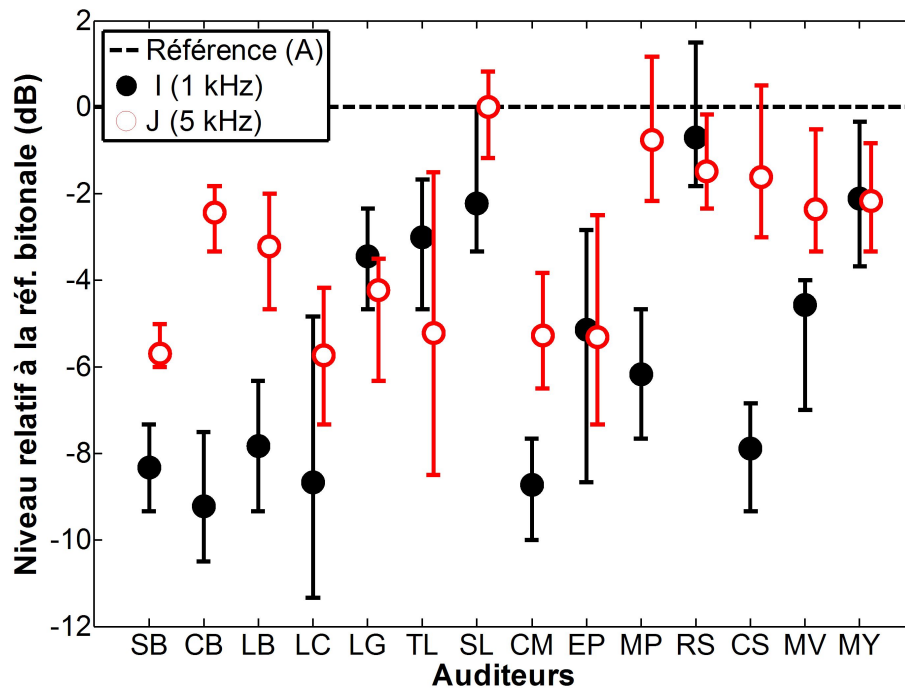


FIGURE 4.7: Sensibilité en sonie pour chacun des 14 auditeurs. Les cercles noirs et rouges représentent respectivement les résultats obtenus pour les stimuli I (1 kHz) et J (5 kHz). Chaque point correspond à la médiane sur les 8 niveaux d'égalisation estimés pour chaque condition expérimentale. Les barres d'erreurs représentent l'intervalle entre les niveaux d'égalisation min. et max. .

différences interindividuelles sont grandes (écarts interquartiles de 5.3 dB à 1 kHz et 3.6 dB à 5 kHz), et la différence entre les sensibilités en sonie obtenues pour I et J n'est pas significative [TDS, N.S.]. Cela concorde avec nos attentes étant donné que les deux sources I et J étaient en théorie calibrée à l'isotonie.

D'autres part, les fréquences centrales étant suffisamment éloignées pour ne pas solliciter les mêmes bandes critiques, la sonie globale de la source bitonale devrait correspondre à la somme des sonies de chaque composante. En admettant que les deux composantes fréquentielles sont à l'isotonie, cela revient à dire que la sonie globale de la source bitonale devrait être égale au double de la sonie de chaque composante tonale. Or, d'après les courbes isosoniques (ISO-226, 2003), doubler la sonie à 60 phones revient à appliquer un gain de 10 dB à 1 kHz et de 10.5 dB à 5 kHz. Les niveaux d'égalisation mesurés pour les stimuli I et J sont donc finalement bien plus faibles que ceux qui étaient attendus.

4.5.1.2 Discussion

Les variabilités intra- et inter-individuelles sont grandes par rapport à ce qui avait été obtenu pour les autres conditions expérimentales. Cela peut traduire en partie la difficulté de la tâche consistant à comparer la sonie de deux sources qui diffèrent à la fois dans le nombre (1 vs. 2) et dans les caractéristiques spectrales (f_1 vs. f_1+f_2 ou f_2 vs. f_1+f_2) des composantes fréquentielles qui les constituent. D'une manière générale, il est établi que des différences intra- et inter-individuelles im-

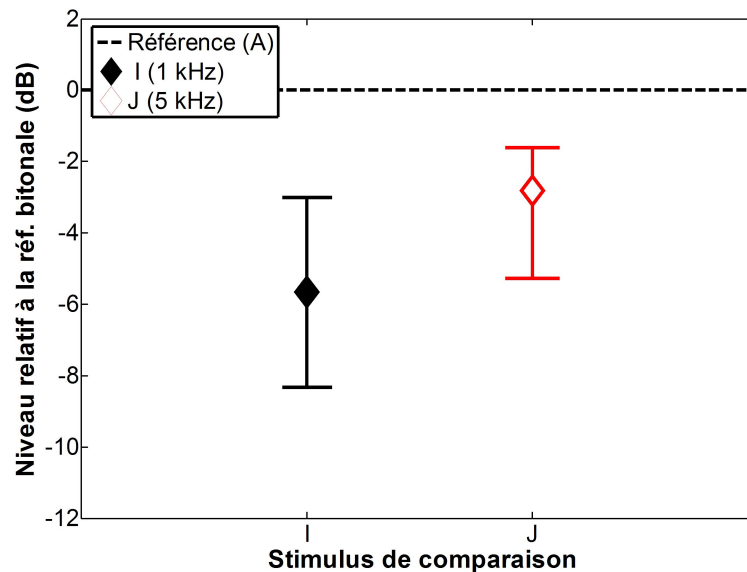


FIGURE 4.8: Sensibilités en sonie médianes en fonction du stimulus de comparaison (I ou J). Les barres d'erreurs représentent les écarts interquartiles associés.

portantes surviennent lorsque la sonie de sources tonales séparées en fréquence est comparée et que cette variabilité dépend de l'ampleur de la séparation fréquentielle (e.g. Robinson et Dadson (1957), Poulsen et Han (2000), Takeshima *et al.* (2001), Takeshima *et al.* (2002)). Par exemple dans l'étude de Takeshima *et al.* (2002) visant à mesurer les courbes de niveau d'isophonie, lorsque la sonie globale d'une source frontale à 1 kHz était comparée à celle d'une source frontale à 5 kHz, les auteurs ont rapporté un écart-type de 6.5 dB à 60 phons, sur un panel de 21 auditeurs. Dans les mêmes conditions, les écarts-types valaient respectivement 5.8 dB et 7.3 dB lorsque la source à 1 kHz était comparée à des sources de fréquence 4 kHz et 6.3 kHz. Le fait de comparer deux sons constitués d'un nombre de composantes fréquentielles différent semble cependant avoir une implication faible sur la variabilité interindividuelle, en comparaison avec l'influence de la séparation en fréquence sur cette même variabilité. L'étude de Florentine *et al.* (1978) rapporte par exemple un écart interquartile d'environ 2 dB (sur un total de 40 jugements par 10 auditeurs) lorsque les auditeurs comparaient la sonie d'un son pur à 1 kHz avec celle d'un son à deux composantes de fréquences 891 Hz et 1122 Hz ($\Delta f = 231$ Hz). Cette variabilité entre les auditeurs est relativement faible. En revanche, pour des stimuli similaires mais avec $\Delta f = 1592$ Hz, les écarts interquartiles devenaient approximativement compris entre 3 et 6 dB, montrant l'importance du rôle joué par l'intervalle fréquentiel (entre les deux composantes du son bitonal et donc entre les deux sons à comparer) sur les différences interindividuelles.

Une étude de Scharf (1969) met également en avant l'évolution des variabilités intra- et inter-individuelles en fonction de l'intervalle fréquentiel entre les deux sons à comparer, dans une configuration expérimentale qui se rapproche de celle de notre expérience. Les auditeurs devaient ajuster la sonie d'un son diotique composé d'un son pur de fréquence f_1 présenté au même niveau à chaque oreille, à la sonie d'un son dichotique composé de deux sons purs de fréquences f_1 et f_2 distribués chacun à une oreille différente. L'étude a montré que la variabilité intra-individuelle augmentait avec

la difficulté de la tâche (à savoir avec l'augmentation de $\Delta f = f_2 - f_1$), la différence entre deux ajustements d'un même stimulus allant de 1 dB en moyenne pour $\Delta f = 0 \text{ Hz}$ à 4 dB pour $\Delta f = 6430 \text{ Hz}$. Dans notre expérience $\Delta f = 4000 \text{ Hz}$ et la différence entre les niveaux d'égalisation minimal et maximal obtenus pour une même condition est en moyenne de 3 dB. Cependant, tandis que dans notre expérience les deux sons à comparer étaient diotiques (incidence frontale), dans l'étude de Scharf (1969) la difficulté de la tâche était probablement plus importante puisque les sons comparés étaient soit un son perçu au centre de la tête, soit deux sons séparés dans chaque oreille.

D'autre part, les niveaux d'égalisation médians mesurés pour les sources tonales I et J au PSE, étaient plus faibles que les niveaux théoriques attendus. Une hypothèse est que la méthode de comparaison impliquait de présenter une des deux composantes seule avant ou après la présentation de la source bitonale, ce qui peut avoir eu pour effet de "capturer" la composante correspondante dans la source bitonale, et mener les auditeurs à grouper les deux composantes de même fréquence.

En effet, ce phénomène a été mis en évidence par exemple dans Bregman et Tougas (1989), où deux composantes tonales (B et C de fréquence respective $f_2 = 848 \text{ Hz}$ et $f_1 = 400 \text{ Hz}$) étaient présentées simultanément, formant ainsi un son complexe. Un son pur (A de fréquence $f_2 = 848 \text{ Hz}$) était ensuite présenté avant ou après les deux composantes (cf. Pattern II et Pattern IV). Les stimuli avaient une durée de 140 ms, étaient séparés par un silence de 30 ms et étaient présentés en boucle (20 répétitions). Dans cette configuration, la frange (A) permettait de capturer la composante de même fréquence B dans le son complexe B-C. La perception des deux sons de fréquence f_2 comme une paire n'était cependant pas évidente, la clarté du percept était évaluée à environ 3/7 par les auditeurs.

Dans Grimault *et al.* (2007) le même schéma est reproduit avec les fréquences $f_1 = 1 \text{ kHz}$ et $f_2 = 2 \text{ kHz}$. Les sons avaient une durée de 200 ms, étaient séparés par un silence de 100 ms et étaient répétés en boucle pendant 45 s. Cependant, la consigne invitait à juger la sonie globale du son complexe, et les auteurs ont effectivement rapporté que le son complexe semblait plutôt intégré par les auditeurs.

L'ajout d'une frange (A) identique à une des composantes (B) du son complexe peut donc permettre de capturer B dans le son complexe B-C, mais le percept n'est pas stable et semble dépendre largement de la consigne donnée à l'auditeur. L'ajout d'une deuxième frange D identique à C permet par exemple d'augmenter la capture de B par A (Bregman et Tougas, 1989). Dans le cadre de notre étude, le son complexe B-C correspondrait à la source bitonale et la frange A à la composante tonale présentée seule avant ou après la source bitonale. Mais ici les sons n'étaient pas répétés en boucle et les échelles de temps étaient plus grandes. L'intervalle de silence entre les deux sons avait une durée de 500 ms. Bien que le percept (intégré ou ségrégué) de la source bitonale puisse être ambigu, il y avait donc peu de probabilité pour que la frange force la ségrégation des composantes de la source bitonale (par capture). Cependant, il était demandé aux auditeurs de comparer la sonie globale des deux sons et nous savons que la difficulté de la tâche augmente avec la séparation en fréquence. Or, l'auditeur essaye toujours de réaliser la tâche la plus simple possible, d'aller vers le percept le plus simple. Ainsi, puisque les composantes pouvaient être ségréguées volontairement (Hirvonen, 2005), et que cela facilitait la tâche de comparaison de sonie, il est possible que l'auditeur se soit finalement servi de la frange pour capturer la composante identique dans le son complexe.

En revanche, ce mécanisme de groupement n'était probablement pas systématique et semblait varier d'un auditeur à l'autre. En terme de sonie, cela impliquerait que la sonie de la composante tonale seule (I ou J) peut être finalement comparée (plus ou moins selon la clarté du groupement) à

la sonie de la même composante dans le son complexe plutôt qu'à la sonie globale de la source bitonale. Un groupement total engendrerait théoriquement un niveau d'égalisation égal à 0 (égalisation de la sonie d'une composante tonale avec elle-même) et inversement, la distinction du son complexe précédé ou suivi d'une composante tonale simple engendrerait en théorie un niveau moyen d'égalisation de sonie sur l'ensemble des auditeurs égal à 10 dB à 1 kHz (théorie de la sommation de sonie). Ce n'est jamais le cas à part pour l'auditeur SL à 5 kHz, ayant produit un niveau d'égalisation nul. D'une manière générale le jugement est donc intermédiaire et l'organisation perceptive semble varier selon l'auditeur et la fréquence. Cela pourrait donc entre autre, expliquer une partie de la variabilité interindividuelle observée dans la figure 4.7.

Enfin, certains auditeurs ont rapporté avoir perçu le son à 5 kHz comme plus "brillant", plus "saillant". Il y a donc peut-être eu dans certains cas, de la gêne introduite pour le son à 5 kHz, amenant les auditeurs à l'ajuster à un niveau d'écoute plus confortable.

4.5.2 Comparaison de sonie entre une source bitonale (f_1+f_2) et une source tonale (f_3)

Dans la partie précédente (§ 4.5.1), la sonie d'une source tonale (f_1) était comparée à la sonie d'une source bitonale ($f_1 + f_2$). Or il s'est avéré que dans le contexte de l'expérience, les auditeurs semblaient égaliser plus ou moins la sonie de la source tonale (f_1) avec la sonie de cette même composante dans le son complexe, au détriment du jugement de sonie globale. Le jugement était la plupart du temps intermédiaire, dépendant vraisemblablement de l'organisation perceptive des différentes composantes mises en jeu. Des mesures supplémentaires ont alors été menées afin de confirmer cette hypothèse et être ainsi en mesure de conclure quant à la validité des modèles pour prédire la sonie globale d'un champ acoustique à deux composantes. La méthode était relativement semblable à la méthode générale, la principale différence étant que la source tonale était centrée à une fréquence f_3 différente de celles des composantes de la source bitonale, afin de forcer l'auditeur à porter son jugement sur la totalité du son complexe.

4.5.2.1 Méthode

Auditeurs Dix des quatorze auditeurs précédemment impliqués dans l'expérience (cinq hommes et cinq femmes âgés de 25 à 50 ans) ont également participé à cette partie supplémentaire.

Stimuli Deux stimuli ont donc été comparés dans cette partie de l'expérience : la source bitonale d'incidence frontale (A) strictement identique à celle utilisée jusqu'à présent (cf. 4.2.3) et une source tonale d'incidence frontale K. Le processus de synthèse de la source tonale K était strictement identique à celui des stimuli I et J, à la seule différence que le bruit à bande étroite utilisé (1/3 d'octave) était cette fois-ci centré à 2236 Hz (fréquence centrale, correspondant à la moyenne géométrique de 1 kHz et 5 kHz).

La calibration pour le stimulus A était la même que dans l'expérience principale. Le stimulus K a quant à lui été calibré à 70 dB SPL en "équivalent champ diffus" (cf. annexe C), puis un gain correctionnel de 0.7 dB était appliqué pour obtenir une équivalence théorique de 70 phones, conformément à l'ISO-226 (2003) et à la correspondance entre champ direct et champ diffus mesurée sur la TA Cortex Mk1 (cf. figure 1.12). Le choix de 70 phones a été réalisé afin d'obtenir une sonie pour la

source tonale proche de la sonie de la source bitonale (compte tenu de l'hypothèse de la sommation de sonie) et ainsi pouvoir faire varier indifféremment l'un ou l'autre des stimuli dans la procédure, les niveaux d'égalisation étant théoriquement égaux au PSE. De plus, cela permet de contraindre la plage de variation des niveaux au cours de la procédure dans un intervalle raisonnable et éviter ainsi les difficultés de mise en place pouvant y être directement liées.

Procédure Les PSEs ont été mesurés à l'aide d'une procédure adaptative à choix forcé sur deux intervalles (2I-2AFC), suivant une règle 1-down/1-up. Chaque essai consistait à présenter successivement un son de comparaison (de niveau variable) et un son de référence (de niveau fixe). Il était demandé à l'auditeur de désigner le son qu'il percevait comme étant le plus fort (1 ou 2).

Ici cependant, dans 50% des conditions le stimulus A était la référence de niveau fixe et le stimulus K était le son de comparaison de niveau variable. Dans l'autre moitié des cas, les rôles des deux stimuli étaient inversés. Les différentes conditions étaient ensuite mélangées dans les différents blocs de mesure. Faire varier soit le niveau de la source tonale soit le niveau de la source bitonale permettait d'éviter de présenter toujours la même référence fixe à chaque essai. Ainsi, cela évitait que l'auditeur finisse par ignorer la référence fixe et base uniquement son jugement sur la comparaison du niveau du son variable avec l'essai précédent (Buus *et al.*, 1997, 1998).

D'autre part, étant donnée la difficulté apparente de la tâche au regard de la variabilité intra-individuelle obtenue pour les conditions I et J, chaque mesure a été répétée six fois au lieu de quatre. Il y avait donc en tout deux conditions correspondant au choix du stimulus que l'on faisait varier (A ou K). Pour chaque condition, deux mesures étaient effectuées avec des niveaux relatifs de départ différents pour le son de comparaison (+10 dB ou -10 dB par rapport au seuil attendu). Au final, l'estimation du PSE pour chaque condition était définie comme la moyenne des 12 niveaux d'égalisation obtenus (2 niveaux de départ \times 6 répétitions de la mesure).

Un total de 24 mesures a été ainsi réalisé par chaque auditeur. Les mesures étaient réparties en 3 blocs de 8 mesures de manière aléatoire, correspondant à une session de test de 45 min.

Le reste de la procédure était strictement identique à celle exposée dans la méthode générale.

4.5.2.2 Résultats

Pour chaque auditeur deux niveaux d'égalisation ont été estimés au PSE : un premier dans le cas où le stimulus K variait (μ_1) et un deuxième dans le cas où le stimulus A variait (μ_2). La différence entre les valeurs absolues de μ_1 et μ_2 était toujours inférieure à 1 dB ; le choix du stimulus variable avait donc peu d'effet sur le niveau relatif d'égalisation. Tous les niveaux d'égalisation ont ensuite été ramenés au cas où le stimulus K variait puis moyennés. Ainsi, les niveaux d'égalisation au PSE finaux étaient estimés par la valeur $\mu = (\mu_1 - \mu_2)/2$. Les sensibilités en sonie représentées sur les courbes correspondent à l'opposé du niveau d'égalisation ($-\mu$).

La figure 4.9 représente les sensibilités en sonie individuelles (cercles blancs) et médiane sur les dix auditeurs (losange noir) pour la source tonale K.

A nouveau, la variabilité interindividuelle est grande, 8.6 dB séparant les seuils obtenus par les auditeurs extrêmes SL et CM. Les sensibilités des autres auditeurs sont réparties de manière relativement homogène sur l'intervalle [-4.6 dB ; +4 dB] défini par les extrêmes.

Quant à elle, la variabilité intra-individuelle est du même ordre de grandeur que celle observée pour les conditions I et J (intervalle entre les seuils min. et max. d'environ 3 dB en moyenne).

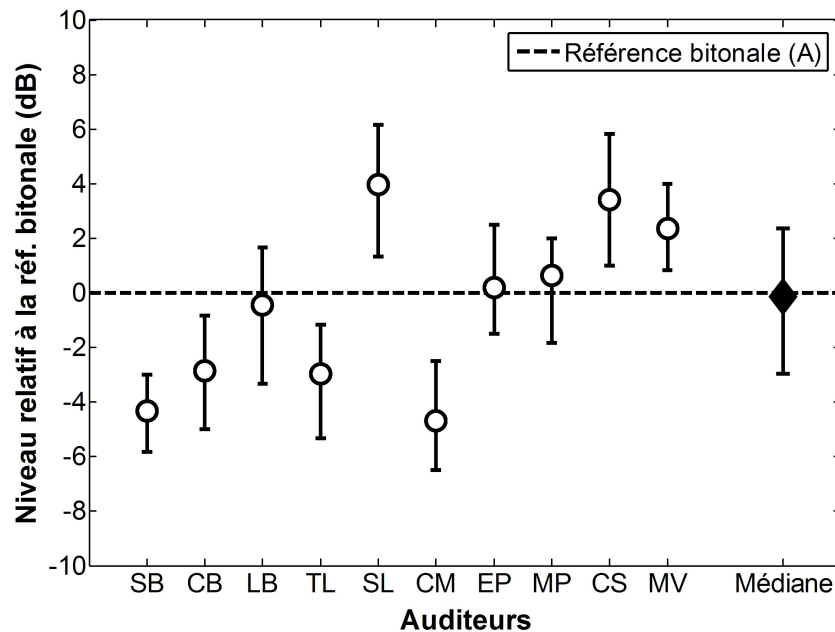


FIGURE 4.9: Sensibilités en sonie individuelles (cercles blancs) et médiane (losange noir) pour la source tonale K par rapport à la référence bitonale A. Les barres d'erreur représentent les valeurs min. et max. sur les différentes répétitions de la mesure pour les données individuelles et l'écart interquartile pour la médiane.

Finalement, malgré la forte variabilité entre les auditeurs, le seuil médian obtenu correspond au seuil théorique attendu. En effet, la sensibilité médiane sur l'ensemble des auditeurs est de 0.1 dB ce qui traduit le fait qu'au PSE, la source tonale à 70 phons a été ajustée au même niveau que la source bitonale composée de deux sources tonales à 60 phons. Ce résultat est donc en accord avec la théorie pour des sons purs, reflétant une simple sommation de la sonie des deux sources tonales à 1 kHz et 5 kHz.

4.5.2.3 Discussion

La sensibilité médiane obtenue pour K (2236 Hz) est en accord avec la théorie, contrairement à ce qui avait été mesuré pour les stimuli I (1 kHz) et J (5 kHz). Cela renforce donc notre hypothèse selon laquelle, pour les stimuli I et J, la composante tonale (I ou J) pouvait être utilisée par l'auditeur pour capturer la composante identique dans la source bitonale et faciliter ainsi la tâche de comparaison en sonie. Selon la clarté du percept, nous pensons que cela amenait les auditeurs à juger plutôt soit la sonie de la composante fréquentielle concernée (1 kHz ou 5 kHz), soit la sonie globale de la source bitonale. En revanche, lorsque la frange se situe exactement entre les deux fréquences du son complexe (stimulus K à 2236 Hz), il n'y aurait plus de réel intérêt à séparer les composantes, car cela ne facilite pas la tâche de comparaison en sonie. D'autre part, la frange ne se situe dans la bande critique d'aucune des composantes, elle ne permet donc plus de les capturer. Cela permettrait ainsi d'expliquer pourquoi les résultats obtenus pour K sont désormais en accord avec la théorie.

Les variabilités intra- et inter-individuelles sont du même ordre de grandeur que celles relevées

pour les conditions I et J. Cela confirme donc d'une part la difficulté de la tâche de comparaison de deux stimuli qui diffèrent à la fois dans le nombre et la fréquence des différentes composantes tonales. D'autre part, le biais variable dans le jugement de sonie globale évoqué plus haut avait une influence sur le niveau d'égaleisation médian, mais cela influait finalement peu sur la variabilité interindividuelle totale. Les variabilités interindividuelles observées ici sont toujours dans l'ordre de grandeur de celles observées dans la littérature pour des tâches plus ou moins similaires (Takeshima *et al.*, 2002; Scharf, 1969).

La figure 4.10 propose une synthèse des résultats médians obtenus pour les stimuli I, J et K. Afin

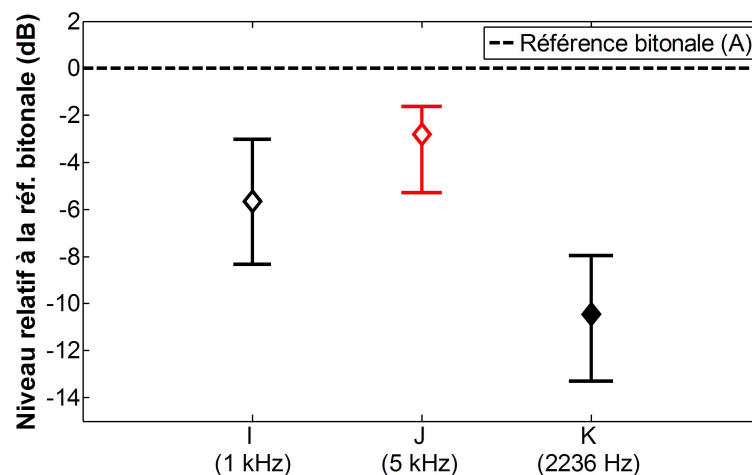


FIGURE 4.10: Comparaison des sensibilités en sonie pour les conditions I, J et K en référence à la source de référence bitonale A. Les marqueurs et barres d'erreurs représentent les médianes sur l'ensemble des 14 auditeurs (pour I et J) et des 10 auditeurs (pour K) ainsi que les écarts interquartiles associés.

de faciliter la comparaison, un gain de -10.3 dB a été appliqué à la sensibilité médiane obtenue pour K. Ce gain correspond au gain à appliquer pour passer de 70 à 60 phones à 2236 Hz en accord avec l'ISO-226 (2003). Le niveau relatif du stimulus K sur la figure est donc en partie théorique et est fourni uniquement à titre indicatif.

La figure permet d'illustrer l'ampleur de l'écart dans le jugement de la sonie globale entre le cas (I et J) où la source tonale de comparaison a une composition fréquentielle en partie identique à celle de la source bitonale et le cas contraire (K) où les spectres n'ont aucun point commun.

4.6 Modélisation

Afin de confronter les modèles de sonie existants aux données expérimentales, les signaux de pression "aux tympans" ont été enregistrés pour tous les stimuli (A à K), strictement dans les mêmes conditions de restitution que celles de l'expérience. Chaque stimulus était au préalable ajusté au PSE pour chaque auditeur (ajustement basé sur les niveaux d'égaleisation mesurés), puis diffusé au casque et enregistré sur la tête artificielle (même TA que pour la mesure des HRTFs et des fonctions de transfert du casque). La sonie a ainsi été estimée par les différents modèles considérés, directement à partir des signaux de pressions reçus aux oreilles de l'auditeur.

4.6.1 Prédiction de la sonie "relative" d'un champ acoustique bitonal

4.6.1.1 Optimisation du gain binaural pour le modèle de Sivonen et Ellermeier (2008)

Dans le chapitre 2, les sensibilités directionnelles en sonie pour une source tonale à 1 kHz ou à 5 kHz avaient été mesurées sur 17 auditeurs. Les données avaient ensuite été modélisées à partir du modèle de sommation binaurale proposé par Sivonen et Ellermeier (2008) en recherchant le gain de sommation binaurale g optimal pour chaque auditeur puis sur l'ensemble des auditeurs. Le gain $g = 3.5 \text{ dB}$ avait finalement été retenu car permettant de représenter au mieux l'ensemble des données (toutes fréquences confondues).

L'ajustement du modèle de sommation par la recherche du gain binaural optimal a été réalisé à nouveau ici, à partir de l'ensemble des données subjectives collectées pour un champ acoustique bitonal (stimuli A, B, C, D, E, F, G et H). Les gains binauraux optimaux ont été estimés indépendamment pour les mesures réalisées sur 14 auditeurs en présence de l'asymétrie dans les fonctions de transfert du casque et pour les mesures répétées sur 6 auditeurs, une fois l'asymétrie corrigée. La méthode de minimisation de l'écart quadratique moyen (EQM) utilisée pour l'optimisation était similaire à la méthode déjà présentée dans la partie 2.3.1, à la seule différence que l'EQM (cf. équation 2.3) entre le modèle et les mesures était calculé directement à partir des valeurs de sonie, pour pouvoir prendre en compte la complexité du spectre :

$$EQM = \frac{1}{n} \sum_{\text{Auditeur}} \sum_{\text{Condition}} \left(\frac{N_{Comp}}{N_{Ref}} - 1 \right)^2 \quad (4.1)$$

n désigne le produit du nombre de conditions (8) par le nombre d'auditeurs (14 ou 6) et N_{Comp}/N_{Ref} est le rapport entre les prédictions de la sonie pour le son de comparaison et pour le son de référence au PSE. La prédiction de la sonie par le modèle de Sivonen et Ellermeier (2008) correspondait à la sonie estimée par le modèle standardisé ISO-532B (1975) à partir des niveaux diotiques équivalents $L_{diot_{eq}}$ (cf. équation 2.2) pour chaque bande de tiers d'octave, ramenés au préalable à des niveaux au centre de la tête (comme le suggéraient les auteurs). Si le rapport N_{Comp}/N_{Ref} vaut toujours 1 au PSE, l'EQM est nul et le modèle est parfait.

Le tableau 4.2 répertorie les gains binauraux estimés pour chaque auditeur et sur l'ensemble des auditeurs ainsi que les valeurs de l'EQM associées. Les gains obtenus pour un champ acoustique bitonal sont présentés dans les deux premières double-colonnes correspondant respectivement à l'expérience principale (sur 14 auditeurs) et à la répétition des mesures sur 6 auditeurs après correction des fonctions de transfert du casque (FTCs). La troisième double-colonne rappelle quant à elle les gains obtenus en présence d'une seule source dans le chapitre 2, sur les 17 auditeurs initiaux.

Les gains estimés pour les six premiers auditeurs (SB, LB, TL, SL, EP et MV) sont pour la majorité similaires entre l'expérience principale (colonne 1) et la répétition de l'expérience (colonne 2), les plus grands écarts étant relevés pour les auditeurs LB (1.2 dB) et SL (1.6 dB). Cela montre une fois de plus que la stratégie de sommation binaurale, directement liée aux DLS mesurées, est relativement stable pour chaque auditeur.

D'autre part, on constate que les gains individuels estimés dans le cas d'un champ acoustique bitonal (colonnes 1 et 2) ont tendance à être toujours légèrement plus élevés que ceux estimés dans un contexte expérimental ne faisant intervenir qu'une seule source tonale (colonne 3). Cette tendance

Auditeurs	Bitonal (1+5 kHz)		Bitonal (1+5 kHz)		Tonal (1 ou 5 kHz)
	14 auditeurs		6 auditeurs		17 auditeurs
	FTCs erronées		FTCs corrigées		(Chapitre 2)
	g	EQM	g	EQM	g
SB	5.6	0.000	5.1	0.000	3.25
LB	3.3	0.001	4.5	0.001	2.89
TL	6.1	0.001	5.5	0.002	5.25
SL	4.6	0.000	3	0.001	3.18
EP	7.8	0.000	9.4	0.001	8.74
MV	3.7	0.000	3.6	0.002	2.11
CB	6.5	0.000			4.47
LC	6.7	0.002			5.76
LG	5	0.000			2.77
CM	5.7	0.000			3.54
MP	5.7	0.001			2.85
RS	3.4	0.002			2.05
CS	5.6	0.000			1.89
MY	5.1	0.000			4.00
FA					2.33
ML					4.48
WA					4.60
Sur tous les auditeurs	5.3	0.002	5	0.003	3.5

Tableau 4.2: Valeurs du gain binaural g optimal obtenues par minimisation de l'EQM et valeurs de l'EQM associées. Les deux double-colonnes de gauche correspondent aux meilleures estimations pour un champ acoustique bitonal (1+5 kHz). La colonne de droite est un rappel des gains obtenus pour un champ acoustique tonal (1 ou 5 kHz). La dernière ligne représente le gain binaural optimal estimé sur l'ensemble des auditeurs.

se retrouve effectivement dans les gains binauraux estimés sur l'ensemble des auditeurs, valant respectivement $g = 5.3 \text{ dB}$ et $g = 5 \text{ dB}$ pour l'expérience principale (colonne 1) et la répétition de l'expérience (colonne 2). Ces valeurs sont très proches (confirmant la stabilité des valeurs estimées) mais plus élevées que la valeur $g = 3.5 \text{ dB}$ issue des mesures de DLS pour une seule source tonale (colonne 3) ou de la valeur $g = 3 \text{ dB}$ suggérée par les auteurs. Il y aurait donc une variation du gain binaural optimal pour le modèle de Sivonen et Ellermeier (2008), en fonction du nombre de sources tonales qui composent le champ sonore.

4.6.1.2 Comparaison des différents modèles

Quatre modèles de sonie ont été comparés dans leurs performances pour prédire la sonie "relative" du son de comparaison par rapport à la sonie de la source bitonale de référence (A) au PSE :

- le modèle de Sivonen et Ellermeier (2008) avec un gain binaural $g = 3 \text{ dB}$, comme suggéré par les auteurs
- le modèle de Sivonen et Ellermeier (2008) avec un gain binaural $g = 5.3 \text{ dB}$ (pour l'expérience principale) ou $g = 5 \text{ dB}$ (pour la répétition de l'expérience) correspondant aux gains optimaux estimés sur l'ensemble des auditeurs à partir des données de cette expérience (cf. Tableau 4.2).
- le modèle standardisé (ANSI-S3.4 (2007))
- le modèle standardisé (ANSI-S3.4 (2007)) avec la prise en compte des inhibitions contralatérales, modélisées suivant les suggestions de Moore et Glasberg (2007).

Les figures 4.11 et 4.12 représentent le rapport entre la sonie prédite pour le son de comparaison (N_{Comp}) et la sonie prédite pour la source bitonale de référence A (N_{Ref}) au PSE, pour chaque condition expérimentale décrivant un champ acoustique bitonal (stimuli A à H). Chaque point représente la valeur médiane du rapport N_{Comp}/N_{Ref} sur l'ensemble des auditeurs (les barres d'erreurs associées représentent les écarts interquartiles).

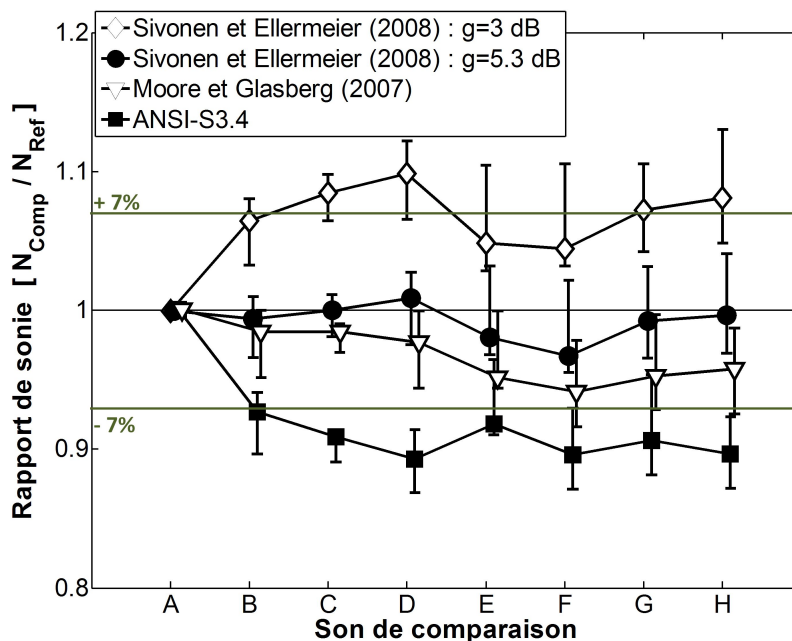


FIGURE 4.11: Prédiction du rapport de sonie au PSE entre le son de comparaison et la source bitonale de référence (A), pour les conditions expérimentales A à H. Prédictions basées sur les résultats de l'expérience principale (FTCs asymétriques).

La figure 4.11 correspond aux prédictions réalisées à partir des mesures sur 14 auditeurs en présence de l'asymétrie dans les fonctions de transfert du casque, tandis que la figure 4.12 correspond aux prédictions réalisées à partir des mesures sur 6 auditeurs, une fois l'asymétrie corrigée (cf. § 4.4.3). On notera que la prédiction des modèles étant directement basée sur les signaux restitués aux oreilles, l'asymétrie dans la correction du casque est bien prise en compte par les modèles et ne devrait donc pas avoir d'influence sur leur performance.

Rappelons également que plus la valeur du rapport est proche de 1, plus les valeurs de sonie estimées par le modèle pour les deux sons sont proches (valeurs normalement égales au PSE) et donc plus le modèle testé est performant. Les valeurs correspondant à une erreur de prédiction de la sonie

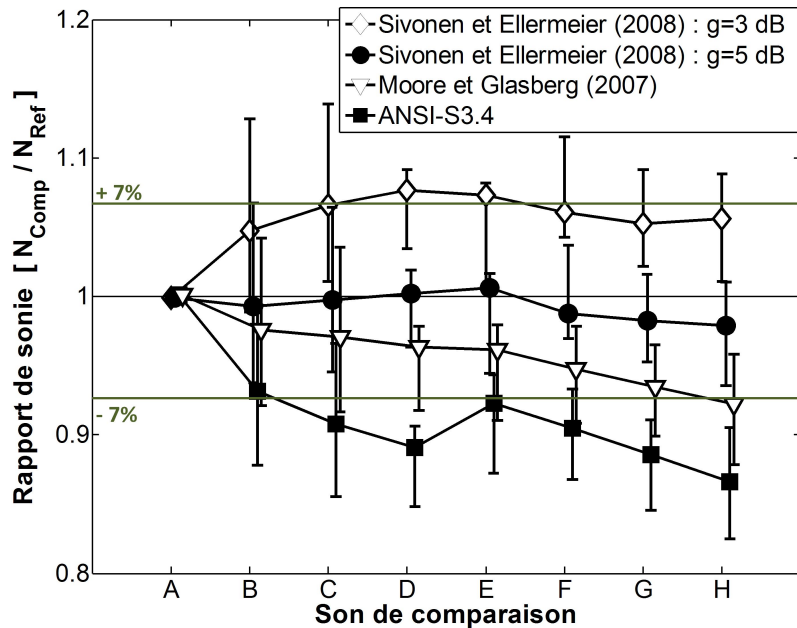


FIGURE 4.12: Prédiction du rapport de sonie au PSE entre le son de comparaison et la source bitonale de référence (A), pour les conditions expérimentales A à H. Prédictions basées sur les résultats de la répétition de l'expérience (FTCs corrigées).

de plus ou moins 7% ont été repérées sur la courbe à titre indicatif. Ce pourcentage correspond de manière symbolique à la variation en sonie engendrée par une différence de niveau de 1 dB (suivant une loi de puissance de la forme $S = k.P^{0.6}$). En-deçà de ce pourcentage d'erreur, les prédictions pourront ainsi être raisonnablement considérées comme suffisamment précises et donc satisfaisantes.

D'après la figure 4.11, l'ensemble des quatre modèles fournit des prédictions dont l'erreur n'excède jamais les 12% pour les rapports médians. Le modèle de Sivonen et Ellermeier (2008) pour $g = 3 \text{ dB}$ a cependant tendance à surestimer légèrement la sonie du son de comparaison, tandis que l'ANSI-S3.4 (2007) a au contraire tendance à la sous-estimer. Les meilleures prédictions sont finalement obtenues par le modèle de Moore et Glasberg (2007) et le modèle de Sivonen et Ellermeier (2008) avec un gain binaural $g = 5.3 \text{ dB}$ optimisé à partir des données. Les mêmes observations peuvent être faites concernant les prédictions des modèles sur les données obtenues en répétant l'expérience après correction des FTCs (figure 4.12). Le modèle de Sivonen et Ellermeier (2008) avec $g = 3 \text{ dB}$ est un peu plus précis que dans le premier cas mais l'ANSI-S3.4 (2007) sous-estime à nouveau la sonie pour l'ensemble des conditions. Les meilleures prédictions sont à nouveau obtenues pour le modèle de Moore et Glasberg (2007) et le modèle de Sivonen et Ellermeier (2008) avec un gain binaural $g = 5 \text{ dB}$ optimisé à partir des données.

Il est intéressant de relever que les sources bitonales B et C avaient été perçues légèrement plus fortes que la source bitonale d'incidence frontale (A), la différence n'étant significative que dans l'expérience principale. On aurait donc pu s'attendre à une sous-estimation de la sonie par les modèles dans la figure 4.11 pour les azimuts de 30° (B) et 60° (C) (par rapport aux azimuts 0° (A) et 90° (D)) et donc à une sous-estimation du ratio N_{Comp}/N_{Ref} dans ces conditions. Or ce n'est pas le cas, l'évolution n'apparaît pas clairement sur les courbes, les modèles ne semblent donc pas sensibles à

cet effet de l'azimut sur la sonie.

Finalement, deux modèles permettent de prédire la sonie "relative" (par rapport à une source bitonale frontale de référence) de manière satisfaisante pour un champ acoustique composé de deux sources tonales simultanées : le modèle de Moore et Glasberg (2007) (ANSI-S3.4 (2007) avec les inhibitions contralatérales) et le modèle de Sivonen et Ellermeier (2008) avec un gain binaural g d'environ 5 dB. Ces modèles permettent notamment de prédire la sonie pour des sources distribuées suivant des grands angles d'incidence (60° et 90°) et donc dans des situations d'écoute fortement dichotiques.

4.6.2 Prédiction de la sonie "absolue" de la source bitonale de référence (A)

Des modèles ont été proposés, permettant de fournir de bonnes prédictions pour un champ acoustique bitonal en considérant la sonie d'une source bitonale d'incidence frontale comme référence. Cependant, pour pouvoir conclure définitivement quant à la validité de ces modèles, il faut également évaluer leur capacité à prédire la sonie de manière "absolue", correspondant à la valeur en sonies fournie. Dans ce but, la sonie fournie par les différents modèles pour la source bitonale de référence a été comparée à la sonie fournie pour une simple source tonale d'incidence frontale (au PSE), permettant de se ramener à une situation diotique simple.

Comme précédemment, la sonie a été estimée à partir des signaux gauche et droit mesurés sur la tête artificielle, suite à la restitution des stimuli préalablement ajustés au PSE pour chaque auditeur. La figure 4.13 représente le rapport entre les prédictions de sonie pour le son de comparaison (I, J ou K) et pour le son de référence (A) au PSE. Chaque point représente la valeur médiane du rapport N_{Comp}/N_{Ref} sur l'ensemble des 14 auditeurs (pour I et J) et des 10 auditeurs (pour K). Les barres d'erreurs correspondent aux écarts interquartiles associés.

Les sources étaient présentées en incidence frontale, la situation d'écoute était donc approximativement diotique. Or, en situation diotique, le choix de la valeur du gain binaural dans le modèle de Sivonen et Ellermeier (2008) n'influe pas sur la valeur du rapport de sonie N_{Comp}/N_{Ref} . De la même manière, lorsque les niveaux aux deux oreilles sont identiques, il n'y a pas d'inhibitions contralatérales (Moore et Glasberg, 2007). Ainsi, seules les prédictions du modèle de Sivonen et Ellermeier (2008) pour $g = 3$ dB et du modèle ANSI-S3.4 (2007) ont été représentées ici.

Concernant les stimuli I et J, l'erreur de prédiction est particulièrement importante, étant comprise entre 25% et 30% pour tous les modèles. Cette erreur illustre à nouveau le fait que les auditeurs ont eu tendance à égaliser en partie la sonie des sources tonales I et J avec la sonie de la composante fréquentielle correspondante (1 kHz ou 5 kHz) dans la source bitonale, au lieu de porter leur jugement uniquement sur la sonie globale. Les niveaux d'égalisation au PSE pour les stimuli I et J étaient alors particulièrement faibles au regard des attentes fondées sur la théorie (cf. figure 4.8). Cela se traduit ici par des valeurs faibles du rapport de sonie N_{Comp}/N_{Ref} au PSE calculées par les différents modèles, mais la performance de ces derniers n'est pas pour autant remise en cause.

Le recours à la source tonale K centrée sur une fréquence intermédiaire a finalement permis d'évaluer plus justement la sonie globale de la source bitonale. Le niveau d'égalisation médian au PSE (sur l'ensemble des auditeurs) était alors en accord avec la théorie (par analogie avec des sons purs). D'après la figure 4.13, il y a bien une amélioration notoire des prédictions pour le stimulus K, mais le rapport de sonie médian N_{Comp}/N_{Ref} reste inférieur à 0.9 pour l'ensemble des modèles. En

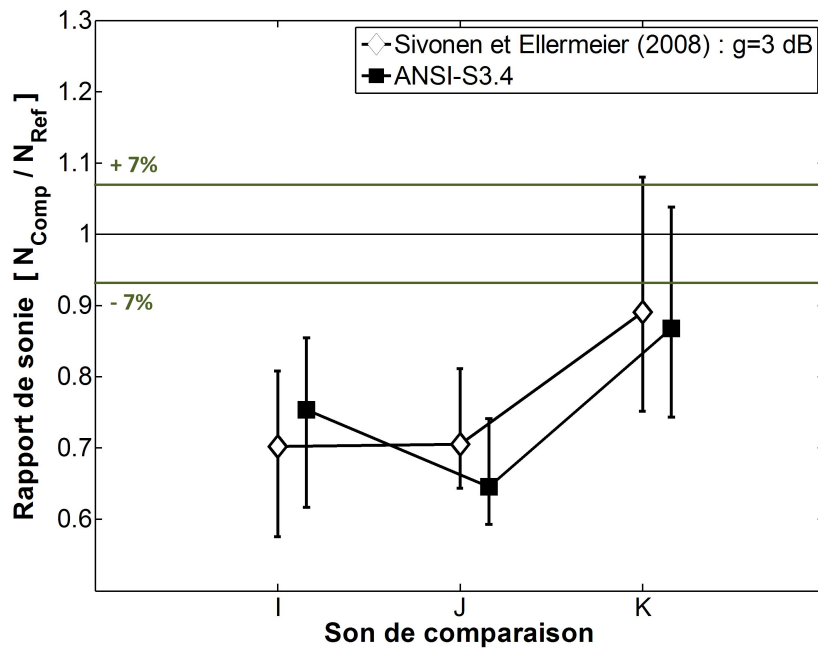


FIGURE 4.13: Prédiction du rapport de sonie médian au PSE entre le son de comparaison et la source bitonale de référence (A), pour les conditions expérimentales I, J et K.

admettant que les prédictions fournies par les différents modèles normalisés pour une source tonale d'incidence frontale en champ libre sont exactes, cela traduirait donc une surestimation systématique de la sonie pour la source bitonale. On notera tout de même que les différences entre les auditeurs sont grandes (reflet de la variabilité déjà constatée pour les niveaux d'égalisation) et que la valeur 1 est comprise dans les écarts-interquartiles.

En terme de sonie "absolue", les valeurs de sonie (en sone) ainsi que les niveaux d'isophonie (en phone) calculés par les différents modèles de sonie au PSE pour le son de référence (A) et les sons de comparaison (I, J et K), sont présentés dans le tableau 4.3.

	A		I		J		K	
	(1+5 kHz)		(1 kHz)		(5 kHz)		(2236 Hz)	
	N	L _N	N	L _N	N	L _N	N	L _N
Sivonen et Ellermeier (g = 3 dB)	11.9	75.7	8.2	70.4	8.0	70	10.6	74.1
ANSI-S3.4	10.3	73.4	7.2	68	6.3	66.1	9.0	71.4
ANSI-S3.4 + Corr. Cortex MK1	13.1	76.9	8.2	70.2	9.1	71.5	11.3	75

Tableau 4.3: Comparaison des valeurs de sonie N (sone) et des niveaux d'isophonie L_N (phone) fournis par les différents modèles, pour les stimuli au PSE.

On constate dans un premier temps que les valeurs en sone fournies par l'ANSI-S3.4 (2007) sont toujours plus faibles que les valeurs fournies par le modèle de Sivonen et Ellermeier (2008) (basé sur l'ISO-532B (1975)), les deux modèles n'étant pas en accord sur la sonie des simples sources tonales.

Une différence importante dans l'implémentation des modèles est la prise en compte des HRTFs (cf. partie 1.4.2.2). Dans le modèle de Sivonen et Ellermeier (2008), l'atténuation appliquée au niveau diotique équivalent ($L_{diot_{eq}}$) pour se ramener d'un niveau "au tympan" à un niveau en champ libre au centre de la tête est basée sur les fonctions de transfert mesurées sur la tête artificielle Cortex Mk1 du laboratoire (pour l'incidence frontale). Or, la fonction de transfert analogue pour le champ libre en incidence frontale prise en compte par la norme ANSI-S3.4 (2007) est bien sûr différente. La dernière ligne du tableau propose donc de prendre en compte la différence entre ces deux fonctions de transfert (représentées dans la figure 1.12 du chapitre 1) afin de pouvoir comparer plus justement les modèles. La correction correspondante, appliquée par bande de tiers d'octave à l'entrée du modèle ANSI-S3.4 (2007), est décrite plus précisément dans la partie 1.4.2.2 et représentée par des barres grises dans la figure 1.12.

Suite à la prise en compte des caractéristiques physiques de la tête artificielle, les deux modèles sont désormais en accord sur la sonie de la source tonale I à 1 kHz (8.2 sones). Cependant, bien que les différences aient été réduites, la sonie diffère toujours pour les sources tonales à 2236 Hz (K) et à 5 kHz (J), correspondant à des différences respectives de niveau d'isophonie de 0.9 phones et 1.5 phones. La source bitonale A contenant notamment une composante à 5 kHz, la sonie estimée pour cette source dépend également du modèle (13.1 sones par le modèle ANSI-S3.4 (2007) avec correction et 11.9 sones par le modèle de Sivonen et Ellermeier (2008), soit une différence de 1.2 phones). Les principales différences pouvant intervenir avant le calcul du modèle ont été prises en compte (HRTFs); les différences qui persistent reflètent donc probablement des différences intrinsèques aux deux modèles standardisés (ANSI-S3.4 (2007) et ISO-532B (1975)) dans le calcul de la sonie. Quelque soit le modèle, la sonie estimée au PSE pour la source bitonale A est cependant toujours plus grande (entre 1.6 et 2 phones) que la sonie estimée pour la source tonale K, suggérant effectivement une surestimation des modèles lorsque le champ acoustique contient deux composantes tonales.

Cela rejoint les résultats de Schlittenlacher *et al.* (2015) mettant en avant une surestimation de la sommation de sonie spectrale par le modèle ANSI-S3.4 (2007) pour des fréquences supérieures à 1 kHz. Dans cette étude, la sonie de bruits large bande (BF : [125 Hz - 1 kHz], MF : [500 Hz - 2 kHz] ou HF : [1.25 kHz - 5 kHz]) était comparée à la sonie d'un son pur à 1 kHz. Les deux modèles testés (DIN-45631 (1990) et ANSI-S3.4 (2007)) surestimaient alors la sonie pour le bruit MF et particulièrement pour le bruit HF, correspondant à la gamme de fréquence des composantes de la source bitonale dans notre expérience. Les prédictions différaient également d'un modèle à l'autre pour le bruit HF, principalement pour les niveaux faibles et modérés. Une surestimation de l'ensemble des modèles (Zwicker *et al.* (1984) ou Moore et Glasberg (1996), Moore *et al.* (1997)) avait également été rapportée dans Meunier *et al.* (2000) pour un bruit centré à 3 kHz, de largeur de bande 2040 Hz à 60 dB SPL. La surestimation des modèles dans notre cas s'explique donc vraisemblablement par une surestimation de la sommation de sonie spectrale des composantes centrées à 1 kHz et 5 kHz, plutôt que comme une conséquence directe du nombre de sources.

Dans les travaux de Song (2007), le modèle de Sivonen ($g = 3$ dB) surestimait la sonie d'environ 1 sone pour les stimuli à deux composantes (centrées à 1 kHz et 3.15 kHz), que les composantes soient focalisées, ou distribuées (à 10 ou 30° d'azimut). La sonie du son de référence était d'environ 5 sones,

ce qui correspondait donc à une surestimation d'environ 20% (sur la moyenne des auditeurs) de la sonie d'un champ acoustique contenant deux composantes. Dans notre étude, une surestimation de la sonie de la source bitonale d'environ 10% à 14% a été observée, confirmant les résultats rapportés par Song (2007). L'ampleur de l'erreur des modèles reste tout de même moins importante et pour un certain nombre d'auditeurs les prédictions sont correctes. Il est probable que pour un plus grand nombre de sources tonales ou pour des sons environnementaux dont l'énergie spectrale est répartie sur l'ensemble du spectre, les modèles surestiment davantage la sonie du champ acoustique.

4.7 Conclusion

La première partie de l'étude consistait à mesurer la sonie directionnelle d'une source bitonale (1 kHz + 5 kHz) à 0, 30, 60 et 90° d'azimut, en s'affranchissant au préalable de la sensibilité directionnelle individuelle en sonie pour chaque composante tonale. Les résultats ont montré qu'un faible effet de l'angle d'incidence sur la sonie persistait, réfutant donc en partie l'hypothèse de départ selon laquelle la sonie directionnelle d'une source multi-tonale pourrait être entièrement expliquée par les sonies directionnelles de chacune de ses composantes tonales. Cependant, l'effet observé est faible et les modèles de sonie ne semblent pas sensibles à cette évolution. D'autre part, la prise en compte des DLS individuelles (mesurées un an plus tôt) pour chaque composante tonale a permis de réduire fortement la variabilité interindividuelle dans les mesures de sonie pour la source bitonale. Cela traduit donc l'importance du rôle joué par les DLS de chaque composante tonale dans la DLS de la source multi-tonale. Mais surtout, cela met en avant la stabilité et la robustesse des DLS mesurées pour chaque auditeur, et donc des stratégies de sommation binaurale individuelles.

La deuxième partie de l'étude consistait à évaluer l'influence de la distribution spatiale de deux sources tonales (1 kHz et 5 kHz) sur la sonie globale, pour des grands angles d'incidence (60° et 90°). La méthode utilisée permettait de s'affranchir entièrement des effets induits par l'angle d'incidence sur la sonie. Les résultats de l'étude principale ont montré qu'il n'y avait pas d'effet de la distribution des sources lorsque la composante à 5 kHz était présentée à gauche. Cependant la sonie des sources distribuées était plus faible lorsque la composante à 5 kHz était présentée à droite, révélant l'existence d'une asymétrie en sonie par rapport au plan sagittal. Des mesures supplémentaires ont révélé l'existence d'une dissymétrie entre l'oreille gauche et l'oreille droite dans les fonctions de transfert du casque, à l'origine d'une différence de niveau à 5 kHz lors de la restitution. Une répétition des deux premières parties de l'expérience sur 6 auditeurs en utilisant des nouvelles fonctions de transfert a permis de montrer que cette différence de niveaux lors de la restitution était effectivement la cause de l'asymétrie observée en sonie. Ainsi, l'ensemble des résultats a permis de conclure d'une manière générale que la distribution spatiale des sources n'avait pas d'effet sur la sonie globale de deux sources tonales simultanées. Ce résultat est une extension pour des grands angles d'incidence (>30°) des résultats déjà rapportés par Song (2007).

Dans une troisième partie, la sonie de la source bitonale de référence (pour toutes les conditions expérimentales) a été comparée à la sonie d'une simple source tonale d'incidence frontale. L'objectif était d'obtenir une validation des modèles pour prédire la sonie globale d'une source à deux composante (précision de la valeur en sones). Dans un premier temps la source bitonale de composition 1+5 kHz était comparée soit à une source tonale centrée à 1 kHz (I), soit à une source tonale centrée à 5 kHz (J). Dans ce cas, les auditeurs avaient tendance à égaliser la sonie des stimuli I et J à des

niveaux faibles par rapport au niveau d'isophonie global théorique de la source bitonale. Il a alors été supposé que certains auditeurs utilisaient le stimulus I ou J présenté avant ou après pour capturer la même composante dans le son complexe (A) afin de faciliter la tâche de comparaison en sonie. Le jugement de sonie globale aurait ainsi été plus ou moins biaisé. L'expérience a ensuite été répétée pour une source tonale centrée à 2236 Hz (K). Malgré l'importante variabilité intra- et inter-individuelle représentative de la difficulté de la tâche, la source tonale à 70 phones a été égalisée au même niveau que la source bitonale contenant deux composantes à 60 phones. Ce résultat en accord avec la théorie pour des sons purs, a permis entre autre de renforcer l'hypothèse avancée pour les stimuli I et J.

Finalement, plusieurs modèles de sonie binaurale ont été testés sur les données subjectives recueillies : 1) le modèle de Sivonen et Ellermeier (2008) avec un gain binaural $g = 3 \text{ dB}$ suggéré par les auteurs, 2) le modèle de Sivonen et Ellermeier (2008) avec $g \simeq 5 \text{ dB}$ correspondant au meilleur ajustement du modèle à partir des données de notre expérience, 3) le modèle standardisé ANSI-S3.4 (2007), 4) l'ANSI-S3.4 (2007) avec la prise en compte des inhibitions contralatérales suggérée par Moore et Glasberg (2007). Tandis que les modèles 1) et 3) avaient tendance respectivement à surestimer ou sous-estimer la sonie du champ acoustique (en référence à la source bitonale frontale), les modèles 2) et 4) permettaient de prédire précisément la sonie "relative" de deux sources tonales, focalisées ou distribuées. La validité des modèles a été mise en évidence pour des sources simulées suivant des grands angles d'incidence ($>30^\circ$), et donc dans des situations d'écoute fortement dichotiques. Cependant, en ce qui concerne la sonie "absolue", l'ensemble des modèles surestimaient la sonie de 10% à 14% lorsque le champ acoustique contenait plusieurs sources tonales. Ce résultat confirme les conclusions similaires de l'étude de Song (2007) à ce sujet. La surestimation de la sommation de sonie spectrale en hautes fréquences par les modèles de sonie fournit une explication plausible de ce phénomène (Schlittenlacher *et al.*, 2015).

Prédiction de la sonie d'un champ acoustique complexe, composé d'une ou plusieurs sources réelles.

Dans le chapitre 4, la sonie d'un champ acoustique complexe a été abordée à travers l'utilisation de deux sources tonales artificielles, spatialisées en champ libre. Cependant, dans la réalité la complexité du champ acoustique peut être due à la complexité du contenu spectral de chaque source, le nombre et l'organisation spatiale des sources qui le constituent ou encore l'acoustique du lieu. Afin de tester la validité des modèles dans des situations plus "naturelles" d'écoute, des mesures de sonie ont ainsi été effectuées pour une source réelle complexe spatialisée suivant différents angles d'incidence (dans une salle de cours) et pour un champ acoustique constitué d'une multitude de sources réelles simultanées (à l'intérieur d'une voiture).

5.1 Méthode

Afin de limiter le temps d'expérimentation par auditeur, l'expérience a été divisée en deux parties. Les parties différaient uniquement dans le choix des stimuli qui les composaient et dans la liste des auditeurs y ayant participé.

5.1.1 Auditeurs

Douze auditeurs parmi le personnel et les étudiants du laboratoire (Institut für Psychologie, Technische Universität, Darmstadt) ont pris part à cette expérience. Onze auditeurs (5 hommes et 6 femmes), âgés de 19 à 57 ans (moyenne d'âge : 25 ans) ont participé à la première partie de l'expérience. Dix auditeurs (3 hommes et 7 femmes), âgés de 19 à 57 ans (moyenne d'âge : 26 ans) ont participé à la seconde partie de l'expérience. Neuf des auditeurs ont finalement participé au test dans son intégralité. Aucun des audiogrammes n'a révélé de pertes auditives (seuils < 20 dB HL entre 125 Hz et 8 kHz), excepté pour l'auditeur 'WE' présentant un seuil de 25 dB HL à 1 kHz pour l'oreille gauche (cf. Annexe A). Les heures de test pour les étudiants étaient soit indemnisées, soit comptabilisées dans le cadre de leur cursus.

5.1.2 Dispositif expérimental

La procédure était implémentée dans Matlab. Les sons étaient générés à l'aide d'une carte son RME Hammerfall DSP Multiface II (16 bits, 44100 Hz) puis restitués à l'auditeur dans une cabine audiométrique à l'aide d'un casque Sennheiser HD 650 (le même que pour les autres expériences

menées au LVA). Un écran était utilisé pour l'affichage, un clavier numérique était utilisé comme terminal de réponse.

5.1.3 Stimuli

Tous les stimuli étaient des enregistrements binauraux de champs acoustiques stationnaires réels, réalisés à l'aide d'une tête artificielle puis restitués au casque. Deux principaux types de champ acoustique ont été distingués : les champs acoustiques "simples" (constitués d'une seule source réelle) et les champs acoustiques "complexes" (constitués de plusieurs sources réelles).

Champs acoustiques "simples" (une source réelle) :

Les champs acoustiques "simples" étaient donc constitués d'une source réelle stationnaire, spatialisée suivant différents angles d'incidence dans une acoustique semi-réverbérante (salle de cours du LVA caractérisée par son volume de $162.4 m^3$ et son temps de réverbération $TR = 1.14 s$). Les principales étapes de la synthèse des stimuli sont schématisées dans la figure 5.1.

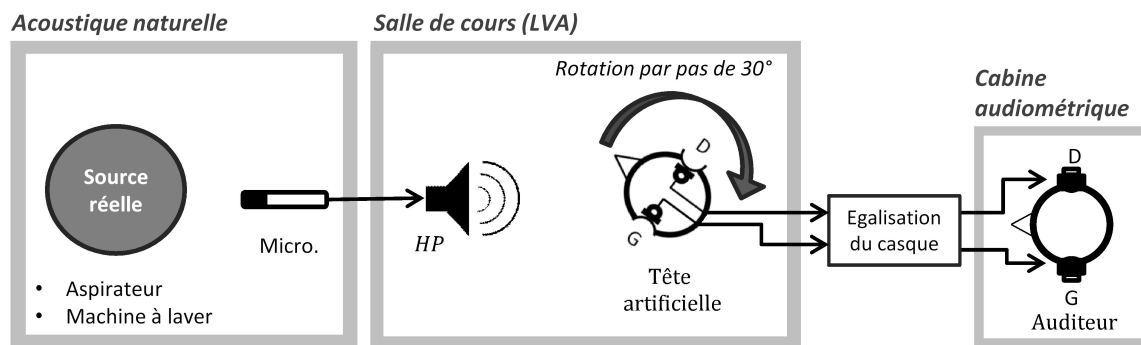


FIGURE 5.1: Schéma des différentes étapes de la synthèse des stimuli, depuis la source jusqu'à la restitution au casque.

Dans un premier temps, un enregistrement monophonique de la source a été réalisé à l'aide d'un microphone placé en face de la source dans son environnement naturel. Puis, le signal enregistré a été restitué sur un haut-parleur dans la salle de cours du LVA et enregistré cette fois-ci en binaural à l'aide d'une tête artificielle Cortex MK1 (sans correction). La tête artificielle était placée sur une plateforme tournante, ajustée pour aligner le centre de la tête avec le centre du haut parleur (à deux mètres de distance). Une rotation de la tête sur elle-même par pas de 30° a ainsi permis de simuler une source provenant de différentes incidences dans le plan horizontal. Les angles d'azimut investigués étaient 0° (incidence frontale), 30° , 60° , 90° , 150° et 180° (source à main gauche). Enfin, une correction des fonctions de transfert du casque¹ (entre 100 Hz et 9 kHz) était appliquée sur les signaux G et D, qui étaient ensuite restitués à l'auditeur au casque dans une cabine audiométrique.

Deux différentes sources sonores considérées comme stationnaires ont été utilisées : un aspirateur (stimuli A1 à A6) et une machine à laver en cours d'essorage (stimuli M1 à M6). Au total, 12 stimuli correspondant à des champs acoustiques "simples" ont été créés (2 sources \times 6 azimuts).

1. Fonctions de transfert du casque présentées dans l'annexe D.2, après correction de l'erreur.

Champs acoustiques "complexes" (plusieurs sources réelles) :

La deuxième partie des stimuli testés correspondait à des champs acoustiques "complexes", dans le sens où ils étaient composés d'une multitude de sources réelles stationnaires réparties dans l'espace.

La création des stimuli consistait simplement à réaliser un enregistrement binaural in situ du champ sonore à l'aide d'une tête artificielle. Les signaux G et D enregistrés étaient ensuite corrigés par les fonctions de transfert du casque (entre 100 Hz et 9 kHz), puis restitués à l'auditeur au casque dans une cabine audiométrique. La figure 5.2 illustre ces différentes étapes.

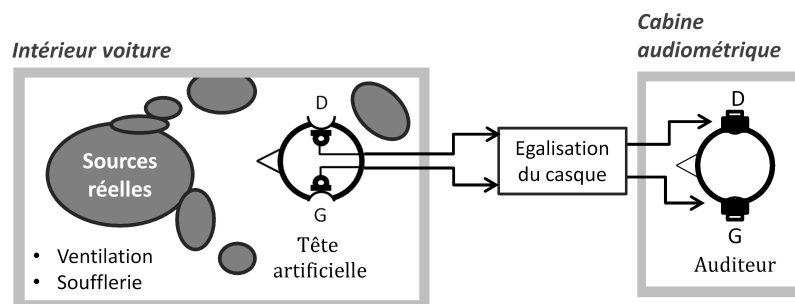


FIGURE 5.2: Schéma du processus de création des stimuli, visant à reproduire au casque le champ acoustique mesuré.

Deux types de sources considérées comme stationnaires ont été testés au cours de cette expérience. D'une part, des bruits de ventilation (air froid, en position dégivrage, sans recyclage d'air) ont été enregistrés à l'intérieur d'une voiture (place conducteur, moteur éteint), à l'aide d'un mannequin acoustique B&K 4100. D'autre part, une voiture a été placée en soufflerie dans une chambre anéchoïque (vitesse simulée à 140 km/h), générant ainsi des sources aéroacoustiques au contact de la carrosserie. Les enregistrements ont été à nouveau réalisés à l'intérieur du véhicule (place conducteur, moteur éteint), à l'aide d'une tête artificielle Head Acoustics HMS III.

Deux enregistrements de ventilation V1 et V2 réalisés respectivement dans une Renault Laguna et une Lancia Lybra ainsi que deux enregistrements S1 et S2 réalisés en soufflerie ont finalement été testés. Les spectres de l'ensemble des stimuli sont représentés dans l'annexe E.

Il est important de souligner que l'égalisation du casque était effectuée à partir de fonctions de transfert du casque mesurées sur la tête artificielle Cortex MK1, tandis que les enregistrements binauraux que l'on souhaitait restituer ici avaient été mesurés sur des têtes artificielles différentes. Ainsi, cela a permis de corriger globalement la couleur spectrale du casque, mais les signaux restitués ne correspondaient pas strictement aux signaux enregistrés, notamment à cause des différences de positionnement des microphones dans le conduit entre les différentes têtes artificielles. Mais cela n'avait pas de réelle importance pour le test, puisqu'il n'y avait pas de référence de localisation pour les stimuli concernés.

Généralités

A chacun des stimuli présentés ci-dessus (A1,...A6, M1,...M6, V1, V2, S1, S2) était associé un son diotique, permettant d'avoir une référence pour les modèles de sonie. Chaque canal du son

diotique était obtenu en moyennant les signaux temporels gauche et droit du stimulus dichotique correspondant ($x_{diot}(t) = (x_G(t) + x_D(t))/2$). Pour les champs acoustiques "simples", le même son diotique était utilisé pour tous les azimuts, calculé à partir des signaux temporels G et D des stimuli d'incidence frontale (A1 et M1).

Les différents stimuli synthétisés sont listés dans le tableau 5.1. L'expérience ayant été scindée en deux parties, les différents stimuli ont été répartis de la manière suivante : l'aspirateur (A1 à A6) et les bruits de ventilation (V1, V2) dans la partie 1, la machine à laver (M1 à M6) et les bruits réalisés en soufflerie (S1, S2) dans la partie 2.

Champ acoustique « simple »						
Source	Aspirateur					
Azimut	0°	30°	60°	90°	150°	180°
Code	A1	A2	A3	A4	A5	A6
Son diotique	$\frac{A1_G(t) + A1_D(t)}{2}$	"	"	"	"	"
Source	Machine à laver					
Azimut	0°	30°	60°	90°	150°	180°
Code	M1	M2	M3	M4	M5	M6
Son diotique	$\frac{M1_G(t) + M1_D(t)}{2}$	"	"	"	"	"
Champ acoustique « complexe »						
Source	Ventilation 1	Ventilation 2	Soufflerie 1	Soufflerie 2		
Code	V1	V2	S1	S2		
Son diotique	$\frac{V1_G(t) + V1_D(t)}{2}$	$\frac{V2_G(t) + V2_D(t)}{2}$	$\frac{S1_G(t) + S1_D(t)}{2}$	$\frac{S2_G(t) + S2_D(t)}{2}$		

Tableau 5.1: Liste des stimuli.

Les spectres par bande de tiers d'octave de tous les stimuli sont représentés dans l'annexe E. Comme pour un grand nombre de sons environnementaux, une grande partie de l'énergie est contenue en basses fréquences pour l'ensemble des stimuli (e.g. Busch-Vishniac *et al.* (2005), Slabbekoorn et Peet (2003)).

D'une manière générale, les sons avaient une durée de 1 s, incluant une croissance et une décroissance de 10 ms (fenêtre de Hann).

Tous les stimuli ont été calibrés à l'aide d'une oreille artificielle B&K type 4153 connectée à un sonomètre B&K type 2250, de manière à ce que le niveau global de restitution du son diotique associé soit égal à 53 dB(A) pour chaque oreille.

5.1.4 Procédure

Les PSEs ont été mesurés à l'aide d'une procédure adaptative à choix forcé sur deux intervalles (2I-2AFC), suivant une règle 1-down/1-up, permettant d'estimer le point à 50% de réponses positives

sur la fonction psychométrique.

Chaque essai consistait à présenter successivement un son de référence de niveau fixe (A1,..., M1,..., V1, V2, S1 ou S2) et un son de comparaison de niveau variable (son diotique associé). L'intervalle inter-stimuli (ISI) était fixé à 500 ms. L'ordre de présentation des deux sons était choisi de manière aléatoire à chaque essai afin d'éviter les erreurs systématiques de type "Time-order errors" (Hellström, 1979). Il était alors demandé à l'auditeur de désigner le son qu'il percevait comme étant le plus fort (1 ou 2). Selon la réponse de l'auditeur, le niveau du son de comparaison était alors augmenté ou diminué d'un pas de variation de 4 dB jusqu'à la 2^e inversion, de 2 dB jusqu'à la 4^e inversion puis de 1 dB jusqu'à la fin de la mesure (10 inversions). Le niveau relatif du son de comparaison par rapport au son de référence était limité dans l'intervalle [-15 dB, +15 dB]. Le niveau d'égalsation était finalement estimé pour chaque mesure en faisant la moyenne arithmétique du niveau relatif du son de comparaison (par rapport au son de référence) sur les six dernières inversions.

Il y avait en tout 16 conditions correspondant aux différents stimuli à tester (A1 à A6, M1 à M6, V1, V2, S1 et S2). Pour chaque condition, deux mesures étaient effectuées avec des niveaux relatifs de départ différents pour le son de comparaison (+10 dB ou -10 dB). Chaque mesure était ensuite répétée deux fois. L'estimation du PSE pour chaque condition, était finalement définie comme la moyenne des 4 niveaux d'égalsation obtenus (2 niveaux de départ \times 2 répétitions). Le test complet comportait ainsi un total de 64 mesures à réaliser.

Comme précisé précédemment, l'expérience était séparée en deux parties. Dans chaque partie, il y avait donc 32 mesures à réaliser, réparties en 4 blocs de 8 mesures de manière aléatoire (les 8 mesures de chaque bloc étant également entrelacées de manière aléatoire).

La durée totale du test était de 2 h (1 h pour chaque partie indépendante).

5.2 Résultats et discussion

Dans un souci de cohérence avec les autres expériences et de clarté dans la présentation des résultats, pour tous les stimuli, le son diotique (qui était le son de niveau variable durant le test) est pris comme référence. Autrement dit, le niveau des sons diotiques est considéré fixe (53 dB(A), cf. 5.1.3), et le niveau des sons dichotiques au PSE (ou leur sensibilité en sonie) est exprimé par rapport au niveau du son diotique pris comme référence. Cela nécessite cependant de faire une approximation en considérant que le niveau auquel auraient été égalisés les différents sons dichotiques pour avoir la même sonie que le son diotique associé, aurait été égal à l'opposé du niveau d'égalsation mesuré dans l'expérience pour le son diotique. Cela n'est pas forcément toujours le cas puisque le PSE n'aurait pas été estimé précisément au même niveau global. L'erreur est cependant supposée relativement faible.

La figure 5.3 présente les sensibilités directionnelles en sonie (individuelles et médianes) obtenues pour l'aspirateur et la machine à laver (A1 à A6 et M1 à M6), par rapport au son diotique pris comme référence (le même pour tous les angles d'azimut).

On constate que la sonie varie avec l'angle d'azimut, de manière très similaire pour l'aspirateur et la machine à laver. Cette similitude est probablement due à l'importance de l'énergie dans les basses fréquences pour les deux sources, partie du spectre sur laquelle l'auditeur devait fonder en grande partie son jugement. En effet, respectivement 70% et 98% de l'énergie spectrale totale était contenue en-dessous de 1 kHz, pour l'aspirateur et pour la machine à laver (cf. spectres, Annexe E).

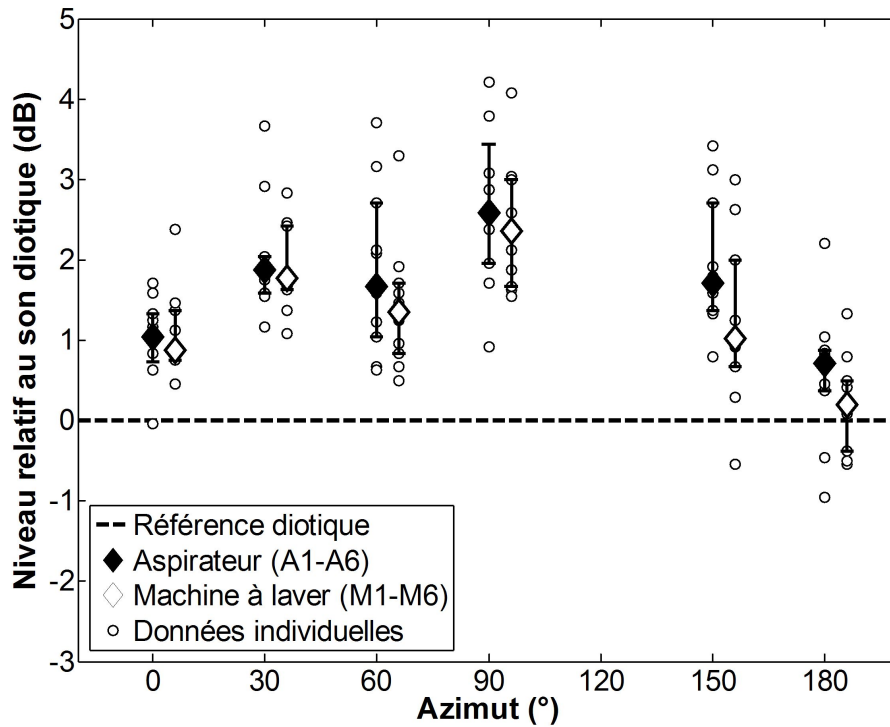


FIGURE 5.3: Sensibilités directionnelles en sonie par rapport au son diotique pris comme référence, pour l'aspirateur et la machine à laver. Les losanges noirs et blancs représentent respectivement la médiane sur 11 auditeurs pour l'aspirateur et la médiane sur 10 auditeurs pour la machine à laver. Les barres d'erreur représentent les écarts inter-quartiles. Les données individuelles sont représentées par des cercles blancs.

A 0° d'azimut, le son dichotique a été perçu environ 1 dB plus fort que le son diotique, pour les deux sources. La sonie du son dichotique atteint son maximum à 90° d'azimut (+2.6 dB pour l'aspirateur et +2.3 dB pour la machine à laver) puis redescend à des valeurs proches de celle du son diotique à 180° (+0.7 dB pour l'aspirateur et +0.2 dB pour la machine à laver). Des ANOVA à mesures répétées sur les données individuelles ont confirmé l'existence d'un effet très significatif de l'angle d'azimut sur la sonie pour l'aspirateur [$F(5, 50) = 13.79$; $p < 0.001$] et pour la machine à laver [$F(5, 45) = 14.12$; $p < 0.001$]. La sonie semble également légèrement plus faible à 60° qu'à 30° d'azimut à la fois pour l'aspirateur et pour la machine à laver. Cependant, les médianes ont été comparées à l'aide d'un test des signes pour chacune des deux sources, montrant que ces différences entre les deux angles n'étaient finalement pas significatives.

L'évolution de la sonie dans le plan horizontal pour ces sources réelles complexes rappelle l'évolution de la sonie mesurée pour une source de bruit à bande étroite (tiers d'octave) centrée à 1 kHz (cf. partie 2.2.3). En effet, dans le plan horizontal la sensibilité directionnelle en sonie variait de manière très similaire, atteignant un maximum de 1.5 dB à 90° d'azimut. La DLS à 60° d'azimut était aussi légèrement plus faible que la DLS à 30°. Ces résultats avaient été obtenus pour une source tonale à 1 kHz, mais il avait été montré dans Sivonen et Ellermeier (2006) que la DLS avait une

évolution d'allure similaire (quoique moins prononcée) à plus basse fréquence (400 Hz). Cependant, les valeurs obtenues ici sont toujours plus élevées d'environ 1 dB, la DLS mesurée pour l'incidence frontale étant notamment différente de zéro. L'existence d'une différence d'énergie entre le son diotique et l'enregistrement binaural pouvant expliquer en partie cette différence est présentée plus loin (page 129).

Concernant la variabilité entre les auditeurs, tous les écarts interquartiles sont inférieurs à 1.5 dB. Cela correspond à l'ordre de grandeur des différences interindividuelles qui avaient été mesurées pour une source tonale à 5 kHz (cf. Chapitre 2), différences plus marquées qu'à 1 kHz. Comme nous l'avions déjà relevé pour les sources tonales, la variabilité interindividuelle augmente également avec les différences interaurales de niveau, les écarts interquartiles étant supérieurs à 1 dB à 60°, 90° et 150° pour l'aspirateur et à 90° et 150° pour la machine à laver. En revanche, ce n'est pas le cas pour la variabilité intra-individuelle qui ne varie pas avec l'azimut. La variabilité interindividuelle reflète donc bien des différences de sensibilité en sonie (donc de stratégie) entre les auditeurs lorsque les différences interaurales sont grandes.

De la même manière pour les champs acoustiques "complexes" (V1, V2, S1 et S2), les sensibilités en sonie individuelles et médianes sont représentées sur la figure 5.4.

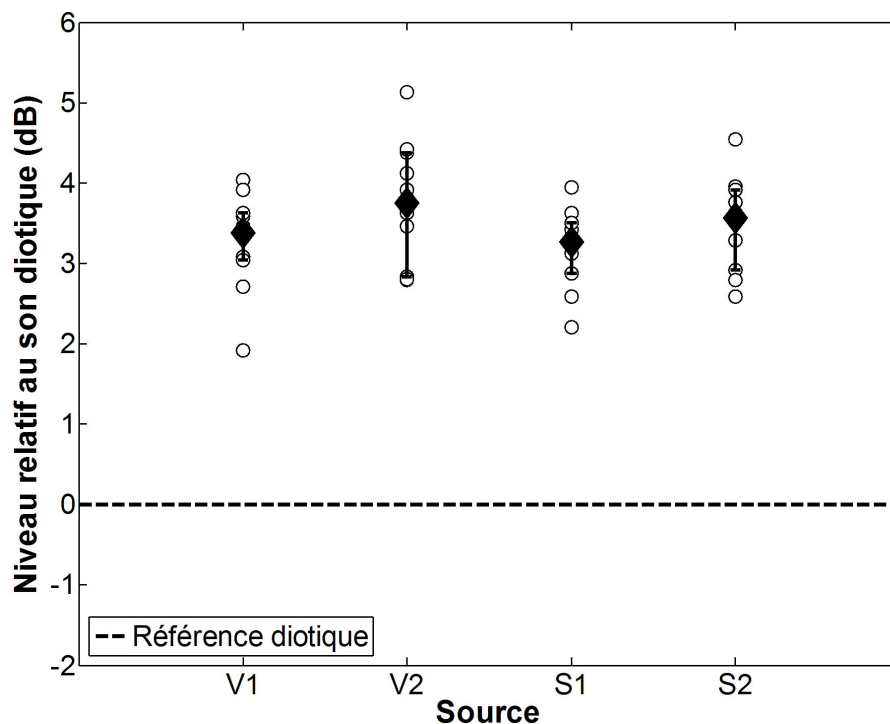


FIGURE 5.4: Sensibilités directionnelles en sonie par rapport au son diotique pris comme référence, pour les stimuli V1, V2, S1 et S2. Les losanges noirs représentent respectivement la médiane sur 11 auditeurs pour V1 et V2 et sur 10 auditeurs pour S1 et S2. Les barres d'erreur représentent les écarts inter-quartiles. Les données individuelles sont représentées par des cercles blancs.

D'après les données médianes, tous les enregistrements binauraux ont été perçus entre 3 dB

et 4 dB plus fort que le son diotique. Ces valeurs sont associées à une variabilité interindividuelle relativement faible, le plus grand écart entre deux auditeurs étant égal à 2.3 dB pour la source V2. L'écart en sonie entre l'enregistrement binaural et le son diotique associé est donc important pour les quatre stimuli testés. Cependant, cette différence peut avoir deux origines différentes : une première d'ordre perceptive (le son dichotique, à énergie égale, serait effectivement perçu plus fort que le son diotique) et une deuxième d'ordre physique (perte d'énergie lors de la synthèse du son diotique).

En effet, lorsque les signaux temporels gauche et droit de l'enregistrement binaural ont été moyennés pour construire le son diotique, les amplitudes des signaux de pression ont été sommées puis divisées par 2. Or, pour que les deux sons aient la même énergie suite à la division des amplitudes par 2 (équivalent à une atténuation de 6 dB), cela nécessite que les signaux gauche et droit soient cohérents (et en phase si on s'attache à une fréquence en particulier). Si les signaux ne sont pas cohérents, ce qui était principalement le cas pour les sons V1, V2, S1 et S2, les signaux se somment en énergie uniquement, ce qui correspond à un gain de 3 dB (au lieu de 6 dB). L'atténuation de 6 dB lors du moyennage n'est donc plus appropriée et entraîne une réduction d'énergie de 3 dB. A titre d'exemple, la figure 5.5 représente les spectres par bandes de tiers d'octave des signaux gauche et droit de l'enregistrement binaural et du son diotique qui en résulte pour le son de ventilation V1. La cohérence entre les signaux est également représentée en parallèle dans la partie basse de la figure.

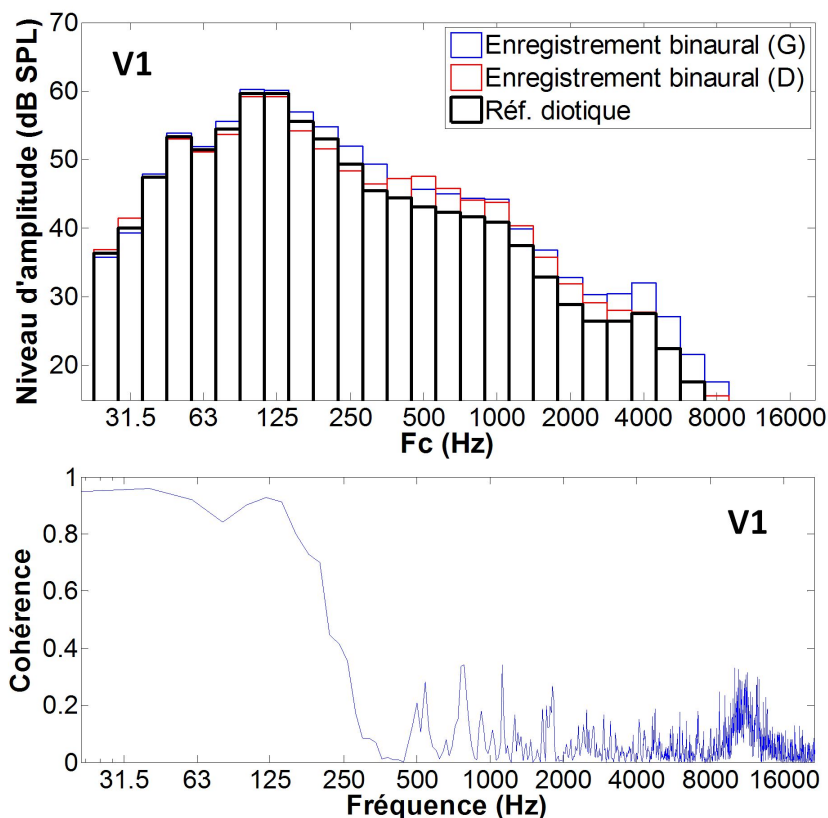


FIGURE 5.5: En haut : Spectres d'amplitude par bandes de tiers d'octave des signaux G et D de l'enregistrement binaural (courbes bleue et rouge) et d'un des canaux du son diotique (courbe noire). En bas : cohérence entre les signaux G et D de l'enregistrement binaural.

On constate bien que, tant que la cohérence est bonne (jusqu'à environ 250 Hz), le niveau du son diotique est compris entre les niveaux G et D de l'enregistrement (le moyennage fonctionne), tandis que lorsque la cohérence chute (au-delà de 250 Hz) le niveau du son diotique chute également, d'environ 3 dB.

Les fonctions de cohérence pour les stimuli V1, V2, S1 et S2 sont représentées dans la figure 5.6 (au milieu et en bas). On constate qu'elles sont relativement similaires. L'atténuation du son diotique introduite lors du moyennage des signaux temporels est donc du même ordre de grandeur pour les quatre stimuli. Une partie de la sensibilité en sonie mesurée (cf. figure 5.4) correspond donc à une différence d'énergie entre l'enregistrement binaural et le son diotique, au-delà de 250 Hz. A titre indicatif, la différence de niveau moyen global (en dB(A)) entre l'enregistrement binaural et le son diotique était respectivement égale à 1.6, 1.9, 2.1 et 2.4 dB pour V1, V2, S1 et S2. En considérant que ces valeurs en dB(A) sont représentatives de l'atténuation réelle du niveau d'isotonie pour le son diotique, la sensibilité en sonie "réelle" serait alors égale à 1.1 dB pour S1 et S2, à 1.7 dB pour V1 et à 1.9 dB pour V2.

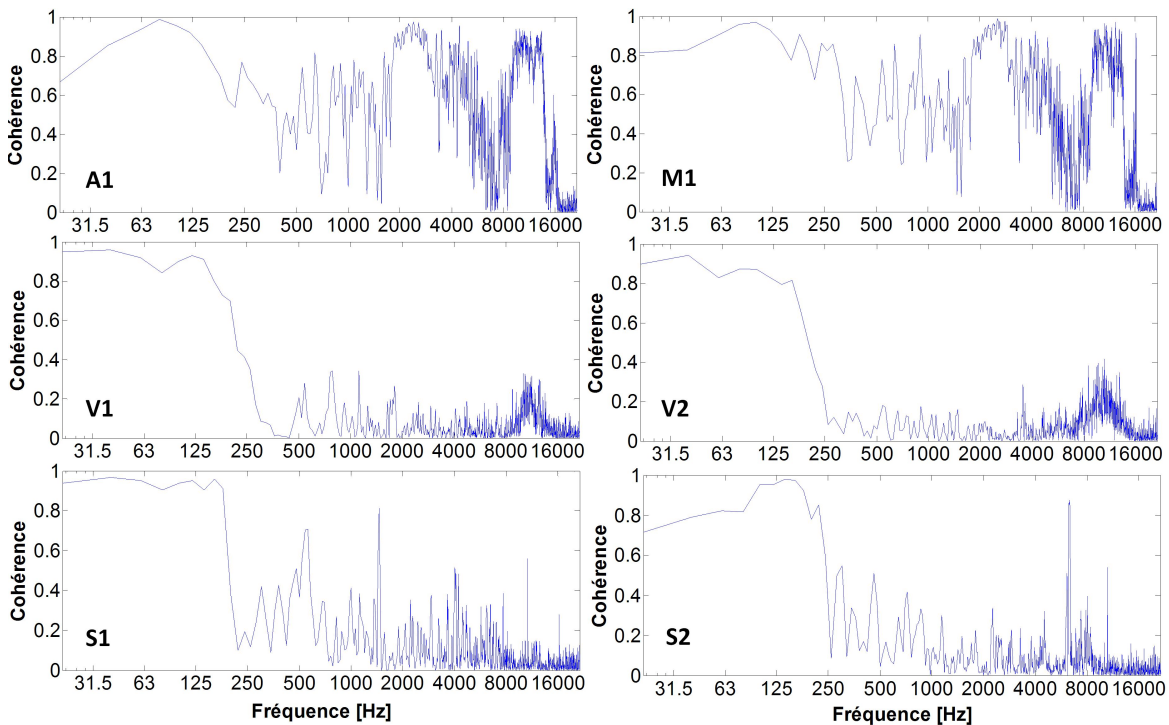


FIGURE 5.6: Fonctions de cohérences entre les signaux gauche et droit pour les stimuli A1 et M1 (en haut), V1 et V2 (au milieu), S1 et S2 (en bas). Pour l'ensemble des stimuli, la cohérence est relativement bonne et stable jusqu'à 250 Hz environ.

Le même phénomène peut également être pris en compte pour l'aspirateur et la machine à laver, intervenant cependant dans une moindre mesure. En effet, les fonctions de cohérences pour les stimuli A1 et M1 sont aussi représentées dans la figure 5.6 (en haut), montrant pour ces stimuli une chute partielle de la cohérence également autour de 250 Hz. Mais les valeurs oscillent ensuite avec la fréquence, atteignant de nouveau des valeurs proches de 1 sur les intervalles [1.8 kHz - 4 kHz]

ou [9 kHz - 13 kHz]. La perte d'énergie lors de la synthèse du son diotique était donc moins importante que pour les stimuli V1, V2, S1 et S2. A titre d'exemple, la différence de niveau moyen global (en dB(A)) entre l'enregistrement binaural et le son diotique était de 0.5 dB pour A1 et de 1.2 dB pour M1. En considérant que ces valeurs sont représentatives de l'atténuation réelle du niveau d'isotonie pour le son diotique, les sensibilités en sonie "réelles" seraient alors égales à 0.5 dB pour A1 et -0.3 dB pour M1. Cela pourrait ainsi expliquer en partie l'écart de 1 dB observé dans les DLS pour l'incidence frontale (cf. figure 5.3). Puisque la même référence diotique était utilisée pour tous les angles d'incidence, cela impliquerait également une surestimation (du même ordre) des DLS "réelles" pour les autres angles d'azimut.

Finalement, une perte d'énergie à partir de 250 Hz a été introduite lors du moyennage des signaux G et D de l'enregistrement binaural pour l'ensemble des stimuli. Cette perte d'énergie n'avait pas de répercussions importantes sur le niveau SPL moyen étant donné que la majorité de l'énergie des signaux est contenue en basse fréquence. Cependant, une atténuation variable du niveau moyen global (en dB(A)) du son diotique a été constatée a posteriori pour l'ensemble des conditions expérimentales, impliquant une réduction potentielle du niveau d'isotonie pour ce son. Ces observations suggèrent que la sensibilité en sonie "réelle" est en réalité plus faible que la sensibilité en sonie "mesurée", notamment pour les stimuli V1, V2, S1 et S2 pour lesquels les signaux G et D de l'enregistrement binaural avaient une cohérence particulièrement faible.

On notera cependant que ces écarts de niveaux n'influent pas sur les prédictions des modèles (§ 5.3), ces dernières étant basées directement sur les signaux de pression enregistrés aux oreilles.

5.3 Validité des modèles

Comme dans les précédentes études, les signaux de pression "aux tympans" ont été mesurés pour l'ensemble des stimuli, restitués au PSE sur la tête artificielle, dans les mêmes conditions que celles de l'expérience.

Les prédictions des différents modèles sont exprimées sous la forme du rapport entre la sonie prédite pour l'enregistrement binaural (N_{Bin}) et la sonie prédite pour le son diotique pris comme référence (N_{Diot}), au PSE. Ainsi rappelons que plus la valeur du rapport est proche de 1, plus les valeurs de sonie estimées par le modèle pour les deux sons sont proches (valeurs normalement égales au PSE) et donc plus le modèle testé est performant.

Les prédictions obtenues à partir des quatre modèles de sonie déjà testés pour les sources artificielles (cf. § 4.6) ont été comparées. Ces modèles incarnaient principalement deux approches de la sommation binaurale : une première approche représentée par le modèle de Sivonen et Ellermeier (2008) avec un gain binaural $g = 3 \text{ dB}$ (suggéré par les auteurs) ou $g = 5 \text{ dB}$ (gain optimal obtenu pour un champ bitonal) et une seconde approche représentée par le modèle ANSI-S3.4 (2007) sans ou avec l'ajout des inhibitions contralatérales suggérées par Moore et Glasberg (2007).

Les prédictions des modèles pour l'aspirateur (A1 à A6), la machine à laver (M1 à M6) et pour les bruits de ventilation ou en soufflerie (V1, V2, S1, S2) sont présentées respectivement dans les figures 5.7, 5.8 et 5.9.

Concernant l'aspirateur (figure 5.7), à 0° d'azimut (stimulus A1) où l'on s'approche d'une situa-

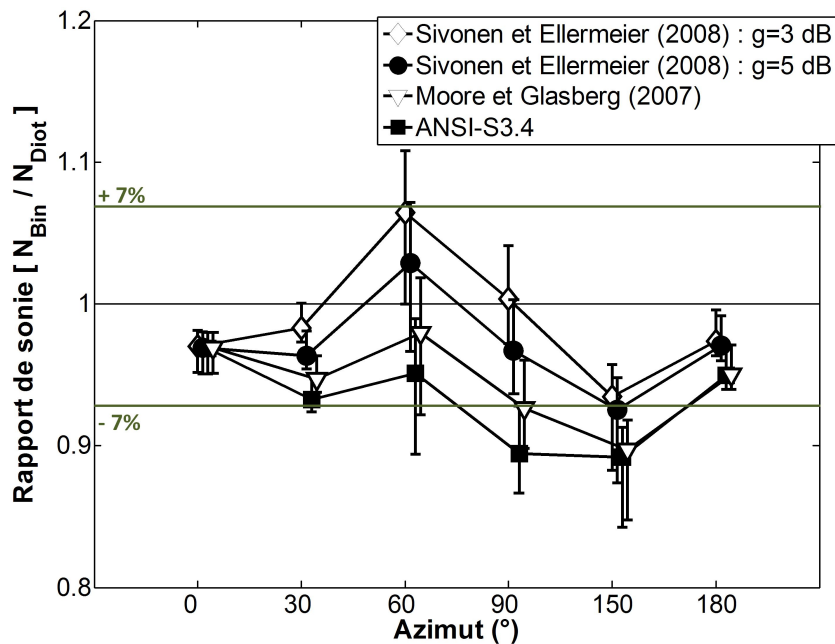


FIGURE 5.7: Prédictions par les différents modèles du rapport de sonie N_{Bin}/N_{Diot} au PSE, pour les stimuli A1 à A6. Chaque marqueur (et barre d'erreur) représente le rapport médian sur 11 auditeurs (et l'écart interquartile associé).

tion d'écoute diotique, tous les modèles sous-estiment la sonie de l'enregistrement binaural de 3%. Cela correspond au fait qu'à énergie égale, l'enregistrement binaural a été perçu légèrement plus fort que le son diotique (de 0.5 dB d'après les estimations en dB(A), cf. § 5.2). La prédiction reste néanmoins très satisfaisante. Pour les autres azimuts (stimuli A2 à A6), toutes les erreurs de prédiction des différents modèles sont inférieures à 7%, à part à 90° et 150° d'azimut (lorsque les ILDs sont grands) où la sonie de l'enregistrement binaural a tendance à être sous-estimée par les modèles basés sur l'ANSI-S3.4 (2007). Les meilleures prédictions sont obtenues par le modèle de Sivonen et Ellermeier (2008) avec $g = 3$ dB ou $g = 5$ dB.

Concernant la machine à laver (figure 5.8), les deux modèles basés sur l'approche de Sivonen et Ellermeier (2008) sont une fois de plus les modèles les plus performants, l'erreur de prédiction étant presque toujours inférieure à 3%. Le modèle de Moore et Glasberg (2007) fournit également des prédictions satisfaisantes pour l'ensemble des angles d'azimut (erreur inférieure à 7%). En revanche, l'ANSI-S3.4 (2007) sans les inhibitions a de nouveau tendance à sous-estimer la sonie, notamment lorsque les différences interaurales de niveau sont grandes. Il est intéressant de souligner également qu'à 0° d'azimut, contrairement à la légère sous-estimation relevée pour les sons d'aspirateur, tous les modèles ont abouti à un rapport de sonie très proche de 1 au PSE.

Enfin, concernant les prédictions pour les champs acoustiques "complexes" (figure 5.9), tous les modèles sous-estiment la sonie de l'enregistrement binaural. Pour les enregistrements réalisés en soufflerie (S1 et S2), la sous-estimation est de l'ordre de 7% pour tous les modèles. L'erreur existe

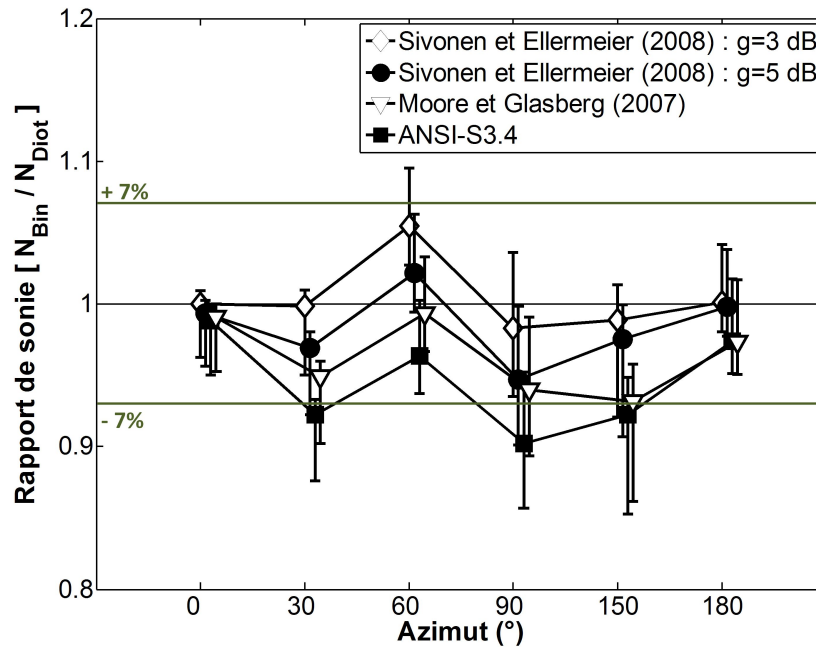


FIGURE 5.8: Prédications par les différents modèles du rapport de sonie N_{Bin}/N_{Diot} au PSE, pour les stimuli M1 à M6. Chaque marqueur (et barre d'erreur) représente le rapport médian sur 10 auditeurs (et l'écart interquartile associé).

mais peut être considérée comme raisonnable. En revanche, pour les sons de ventilation (V1 et V2), la sous-estimation par les modèles basés sur l'ANSI-S3.4 (2007) est comprise entre 8 et 10%, et atteint 11 à 14% pour le modèle de Sivonen et Ellermeier (2008). Tous les modèles sont ainsi mis en échec.

Une hypothèse pour expliquer cette sous-estimation des modèles de sonie en présence de champs acoustiques "complexes" concerne l'effet de la corrélation interaurale sur la sonie. En effet, il a été montré récemment que la manipulation de la corrélation interaurale sur l'ensemble du spectre des signaux gauche et droit pouvait avoir une influence sur la sonie globale (Edmonds et Culling, 2009) et donc sur la performance des modèles (Miranda et Cabrera, 2008). Bien que cela contredise les résultats d'études plus anciennes réalisées sur des bruits large-bande (Dubrovskii *et al.*, 1972; Eichenlaub *et al.*, 1996), les bruits avec une corrélation interaurale nulle ($\rho = 0$) ont été perçus environ 2 dB plus fort que les bruits corrélés (qu'ils soient en phase ($\rho = 1$) ou en opposition de phase ($\rho = -1$)) dans le cas où la majorité de l'énergie était contenue en basses fréquences. Ce phénomène a été observé par les auteurs pour des bandes de bruit de largeur 1 ERB, centrées à des fréquences de 250, 500, 1000 et 2000 Hz. Cependant pour la fréquence centrale 4000 Hz, la différence de niveau au PSE entre le cas décorrélié ($\rho = 0$) et les cas corrélés ($\rho = -1$ ou $\rho = 1$) a été réduite à 1.3 dB (au lieu de 2 dB pour les fréquences plus basses). L'effet était également fortement réduit pour un bruit de bande plus large à des fréquences plus élevées (2894-4763 Hz) et inexistant pour un bruit large bande (100-5000 Hz).

Dans notre cas, les stimuli V1, V2, S1 et S2 contenaient la majorité de leur énergie dans le bas du spectre (entre 90% et 98% de l'énergie était contenue en-dessous de 500 Hz, cf. Annexe E). De plus,

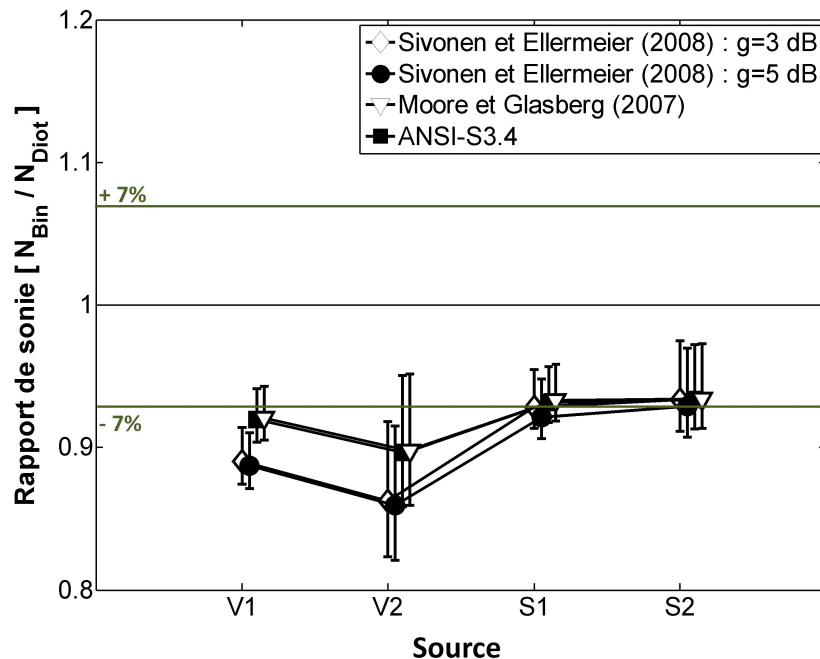


FIGURE 5.9: Prédictions par les différents modèles du rapport de sonie N_{Bin}/N_{Diot} au PSE, pour les stimuli V1, V2, S1 et S2. Chaque marqueur (et barre d'erreur) représente le rapport médian sur 10 (ou 11) auditeurs (et l'écart interquartile associé).

bien que la fonction de cohérence entre les signaux gauche et droit des enregistrements binauraux ait eu une valeur proche de 1 pour les très basses fréquences, elle chutait systématiquement à une valeur proche de zéro à partir de 250 Hz (cf. figure 5.6). Il est donc possible d'après l'étude de Edmonds et Culling (2009), que les enregistrements binauraux (avec une faible cohérence) aient été perçus plus fort que les signaux diotiques (parfaitement cohérents), l'effet de la cohérence des signaux sur la sonie n'étant pas prise en compte dans les modèles existants. La plus grande différence de niveau mesurée par les auteurs à l'isophonie (entre $\rho = 0$ et $\rho = -1$ ou $\rho = 1$) était égale à 2 dB (Edmonds et Culling, 2009). Pour une fonction de sonie de la forme $S = k \cdot P^{0.6}$, cela représente schématiquement une variation de sonie de 14%, ce qui coïncide avec l'ordre de grandeur des pourcentages d'erreurs relevés dans la prédiction des modèles. Une sous-estimation importante de tous les modèles avait également été observée pour un bruit rose dans Miranda et Cabrera (2008) lorsque la corrélation interaurale était faible ; principalement lorsque l'IACC (Interaural Cross Correlation) était inférieur à 0.3 (d'après la figure 6 de Miranda et Cabrera (2008)).

Pour les champs acoustiques "simples" (stimuli A1 à A6 et M1 à M6), la cohérence était globalement meilleure sur l'ensemble du spectre, ce qui expliquerait pourquoi les modèles sous-estiment la sonie principalement pour les champs acoustique "complexes". Cependant, pour l'aspirateur en incidence frontale, tous les modèles avaient tendance à sous-estimer légèrement la sonie également, ce qui n'était pas le cas pour la machine à laver. Il se trouve que les sons de machine à laver contenaient d'avantage d'énergie en-dessous 250 Hz que les sons d'aspirateur (cf. Annexe E), région où la cohérence des signaux était bonne et stable (cf. figure 5.6). En revanche, les sons d'aspirateur contenaient

plus d'énergie que les sons de machine à laver au-delà de 250 Hz, là où la cohérence pouvait varier de manière importante. Cela pourrait donc justifier une légère augmentation de la sonie globale pour les stimuli A1 à A6, qui serait très faible ou inexistante pour les stimuli M1 à M6. Cette augmentation de la sonie se traduirait donc par un rapport N_{Bin}/N_{Diot} inférieur à 1 à 0° d'azimut, mais également par des valeurs du rapport de sonie légèrement plus faibles pour les autres angles d'azimut. Cela pourrait par exemple expliquer en partie la sous-estimation particulièrement importante des modèles à 150° d'azimut pour l'aspirateur.

D'une manière générale, la corrélation interaurale n'est pas prise en compte dans les modèles de sonie existants, mais pourrait bien être la cause d'une augmentation de la sonie globale et donc d'une sous-estimation des modèles lorsque les signaux sont décorrélés. Dans notre cas, cela permettrait d'expliquer la mise en échec des modèles pour les champs acoustiques "complexes" et dans une moindre mesure, la faible sous-estimation des modèles dans le cas de l'aspirateur.

5.4 Conclusion

La sonie a été mesurée pour des champs acoustiques "simples" (constitués d'une seule source réelle spatialisée suivant différents angles d'incidence) et pour des champs acoustiques "complexes" (composés d'une multitude de sources réelles), par comparaison avec la sonie d'une référence diotique de contenu spectral analogue. Différents modèles de sonie binaurale ont ensuite été testés sur les données afin de tester leurs performances pour prédire la sonie d'une ou plusieurs sources acoustiques réelles.

Concernant les champs acoustiques "simples", les modèles basés sur l'approche de Sivonen et Ellermeier (2008) avec $g = 3 \text{ dB}$ ou $g = 5 \text{ dB}$ fournissent les prédictions les plus précises sur l'ensemble des conditions expérimentales. Le modèle de Moore et Glasberg (2007) fournit également des prédictions satisfaisantes (erreur inférieure à 7% dans la quasi-totalité des conditions), mais tend à sous-estimer légèrement la sonie par rapport au modèle de Sivonen et Ellermeier (2008), notamment lorsque les différences interaurales sont grandes. En revanche, pour les champs acoustiques "complexes", tous les modèles sous-estiment la sonie (entre 8% et 14%) pour les quatre stimuli testés (V1, V2, S1 et S2). Une hypothèse a été avancée pour expliquer cette sous-estimation, basée sur l'influence de la corrélation interaurale sur la sonie globale (Edmonds et Culling (2009), Miranda et Cabrera (2008)).

Conclusion générale

Les modèles de sonie normalisés (ISO-532B (1975), ANSI-S3.4 (2007), DIN-45631 (1990)) permettent de prédire la sonie dans des situations d'écoutes monaurales (présentation à une seule oreille) ou diotiques (correspondant à une source en incidence frontale), en champ libre, ou en champ diffus. En revanche, les modèles échouent pour prédire la sonie dans des situations "naturelles" d'écoute (différences interaurales, sources non-stationnaires, multiplicité des sources,...). Dans ce contexte, le travail de thèse a été centré sur l'étude de la sonie binaurale des champs acoustiques stationnaires, dans des situations d'écoute réalistes, fortement dichotiques. Les champs acoustiques étudiés impliquaient une ou plusieurs sources, artificielles ou réelles, restituées au casque par un processus de synthèse binaurale générique pour les sources artificielles ou par simple restitution binaurale pour les sources réelles.

En admettant l'hypothèse selon laquelle la sonie binaurale peut être prédite uniquement à partir des signaux reçus aux oreilles de l'auditeur, l'étude avait notamment pour but d'apporter de nouveaux éléments dans la compréhension des phénomènes permettant à l'auditeur de former un unique percept de sonie à partir des signaux reçus par chaque oreille. Différents aspects de la sonie binaurale en présence d'une ou deux sources tonales spatialisées ont ainsi pu être abordés, dont la variabilité interindividuelle et la robustesse de la stratégie de sommation binaurale. Ces différents aspects sont détaillés dans les paragraphes ci-dessous, à travers les résultats expérimentaux présentés dans les chapitres 2, 3, 4 et 5. Un autre objectif important de l'étude était d'évaluer la performance des modèles de sonie binaurale existants (ANSI-S3.4 (2007), Moore et Glasberg (2007), Sivonen et Ellermeier (2008)), pour prédire la sonie dans des situations d'écoute fortement dichotiques, et réalistes. Les conclusions à ce sujet sont résumées dans le dernier paragraphe.

Sommation binaurale pour une source tonale en champ libre La reproduction des mesures de sonie directionnelle réalisées dans les travaux de Sivonen et Ellermeier (2008) pour une source de bruit à bande étroite, sur un plus grand nombre d'auditeurs (17) et pour une plus grande variété d'angles d'incidence (chapitre 2), a permis de confirmer les conclusions déjà rapportées par ces auteurs. La sonie binaurale varie avec l'angle d'incidence de la source et l'amplitude de variation augmente avec la fréquence (plage de variation de la sensibilité directionnelle en sonie (DLS) dans le plan horizontal de 3 dB à 1 kHz et de 6.5 dB à 5 kHz). De plus, un effet de l'angle d'élévation sur la sonie a été mis en évidence : la variation de sonie est plus importante dans le plan horizontal que dans le plan à 45° d'élévation. Pour les données médianes, l'ensemble des variations de la sonie globale avec l'angle d'incidence peut être décrit précisément à partir de la variation des niveaux de pression aux oreilles. La sommation binaurale peut en effet être correctement décrite par l'équation 5.1, où L_{mon} désigne le niveau monaural équivalent (produisant la même sonie), L_G et L_D sont les niveaux aux oreilles, et g est le gain de sommation binaurale.

$$L_{mon} = g \cdot \log_2 \left[2^{\frac{L_G(B_N)}{g}} + 2^{\frac{L_D(B_N)}{g}} \right] \quad (5.1)$$

La valeur $g = 3.5 \text{ dB}$ optimale pour prédire les données médiane, rejoint la valeur $g = 3 \text{ dB}$ suggérée par Sivonen et Ellermeier (2008).

Variabilité interindividuelle Bien que le modèle de sommation binaurale décrit dans l'équation 5.1 fournisse une prédiction fiable de la sensibilité directionnelle en sonie médiane sur l'ensemble des auditeurs, il existe des différences interindividuelles importantes. Le chapitre 2 a permis de quantifier la variabilité interindividuelle dans les mesures de sonie directionnelle (en accord avec les résultats de Sivonen et Ellermeier (2008)) et de montrer que les différences de DLS entre les auditeurs correspondaient à des stratégies de sommation binaurale différentes, relativement stables sur l'ensemble des conditions expérimentales. Les différences de stratégies se traduisaient notamment par des valeurs du gain de sommation binaurale différentes.

Dans un deuxième temps (chapitre 3), les exposants des fonctions de sonie monaurales ont été estimés pour quatre auditeurs ayant participé à l'étude du chapitre 2. Les résultats ont montré que dans une situation fortement dichotique d'écoute, les différences interindividuelles d'exposant pouvaient entraîner des différences interindividuelles importantes dans le rapport de sonie entre les deux oreilles et donc dans les DLS (jusqu'à 2.4 dB d'écart dans le cas d'une sommation de sonie parfaite). Cependant, les différences de DLS ainsi prédites entre deux auditeurs sont parfois contradictoires avec les différences de DLS mesurées. Dans un cas idéal où il n'y aurait pas de contradiction, la variabilité des DLS prédites ne pourrait pas expliquer plus de 50% des DLS mesurées. Dans tous les cas, d'autres facteurs plus importants (probablement d'ordres cognitifs) interviennent dans les stratégies individuelles de sommation binaurale et sont à l'origine de la plus grande part de la variabilité en sonie observée entre les auditeurs.

Robustesse de la sommation binaurale individuelle Dans le chapitre 4, une expérience a consisté à évaluer la sonie directionnelle d'une source bitonale (1 kHz + 5 kHz) dans le plan horizontal, en s'affranchissant au préalable des DLS individuelles mesurées pour chacune des composantes tonales (chapitre 2). Un effet faible (mais significatif) de l'angle d'incidence sur la sonie a été observé, réfutant en partie l'hypothèse selon laquelle la DLS d'une source multi-tonale peut être expliquée par les DLS de chacune de ses composantes tonales. En revanche, la prise en compte des DLS individuelles (un an après leur mesure dans le chapitre 2) a permis de réduire fortement la variabilité interindividuelle dans les mesures de sonie directionnelle pour la source bitonale. Ce résultat a ainsi permis de mettre en avant la robustesse des DLS individuelles, et donc des stratégies individuelles de sommation binaurale.

Sommation binaurale en présence de deux sources tonales en champ libre Dans le chapitre 4, la première partie de l'expérience a mis en avant l'existence d'un effet faible mais significatif de l'angle d'azimut sur la sonie, propre à la source bitonale (1 kHz + 5 kHz). La deuxième partie a permis quant à elle de montrer que la distribution spatiale des deux composantes tonales (1 kHz et 5 kHz) n'avait pas d'influence sur la sonie globale, pour des grands angles d'incidence (60° et 90°). Finalement, comme dans le cas d'une unique source tonale en champ libre, la sommation binaurale

peut être correctement décrite par la combinaison des niveaux de pression suggérée dans l'équation 5.1, pour chaque bande de tiers d'octave. En revanche, en présence de deux sources tonales, la valeur optimale du gain binaural pour prédire les données médianes était $g \simeq 5 \text{ dB}$.

Performance des modèles de sonie binaurale Les modèles de sonie binaurale existants (ANSI-S3.4 (2007), Moore et Glasberg (2007), Sivonen et Ellermeier (2008)) ont pu être confrontés à l'ensemble des données expérimentales, dans des situations d'écoute dichotiques impliquant différents niveaux de réalisme.

Dans le cas de sources artificielles spatialisées dans un environnement réel (chapitres 2 et 4), les modèles de Moore et Glasberg (2007) et Sivonen et Ellermeier (2008) ont permis de prédire la sonie "relative" (par rapport à la sonie d'une référence bitonale d'incidence frontale)) de manière raisonnablement précise. Dans le modèle de Sivonen et Ellermeier (2008), la valeur du gain binaural g a été optimisée pour prédire au mieux les données médianes sur l'ensemble des auditeurs. Les valeurs du gain binaural finalement retenues étaient $g = 3 \text{ dB}$ dans le cas d'une seule source tonale spatialisée et $g = 5 \text{ dB}$ en présence de deux sources tonales spatialisées. Cependant, l'ensemble des modèles surestimaient la sonie globale de 10 à 14% en présence de plusieurs sources tonales, en accord avec les conclusions de Song (2007) à ce sujet. Une explication a été proposée, basée sur la surestimation de la sommation de sonie spectrale par les modèles au-delà de 1 kHz (Schlittenlacher *et al.* (2015)).

Dans le cas de sources réelles (chapitre 5), la sonie a été mesurée pour des champs acoustiques "simples" (constitués d'une seule source réelle spatialisée suivant différents angles d'incidence) et pour des champs acoustiques "complexes" (composés d'une multitude de sources réelles), par comparaison avec la sonie d'une référence diotique de contenu spectral analogue. Pour les champs acoustiques "simples", le modèle de Sivonen et Ellermeier (2008) fournissait des prédictions satisfaisantes pour l'ensemble des stimuli, quelque soit le gain binaural utilisé (3 ou 5 dB). Malgré sa tendance à sous-estimer légèrement la sonie lorsque les différences interaurales étaient importantes, le modèle de Moore et Glasberg (2007) a également fourni des prédictions raisonnables. En revanche, tous les modèles échouaient pour prédire la sonie des champs acoustiques "complexes" (sous-estimation de 8 à 14%). Une hypothèse fondée sur l'influence de la corrélation interaurale (Edmonds et Culling, 2009) a été avancée pour expliquer cette sous-estimation (cf. *Perspectives*).

D'une manière générale, dans l'ensemble des situations dichotiques d'écoute, le modèle standard ANSI-S3.4 (2007) tendait à sous-estimer la sonie globale.

Perspectives

Ce travail a ouvert plusieurs perspectives de recherche concernant la sonie binaurale. Dans un contexte où les modèles de sonie basent leurs calculs uniquement sur les spectres d'amplitude des signaux de pression reçus aux oreilles de l'auditeur, les principales perspectives concernent l'étude et la prise en compte d'autres paramètres physiques extraits des signaux acoustiques (ITD, corrélation interaurale, déphasage) d'une part et de certains facteurs sensoriels et cognitifs indépendants des signaux acoustiques (ASA, constance perceptuelle,...) d'autre part. Quelques suggestions sont précisées ci-dessous.

Dans le chapitre 5, l'ensemble des modèles de sonie binaurale ont sous-estimé la sonie des champs acoustiques "complexes" constitués d'une multitude de sources réelles. Ces stimuli contenant la majorité de leur énergie en basse fréquence, ont été caractérisés par une corrélation interaurale très faible au-delà de 250 Hz. Or il a été récemment mis en évidence par Edmonds et Culling (2009) que des bruits présentés avec une corrélation interaurale nulle étaient perçus plus fort que des bruits corrélés (qu'ils soient en phase ou en opposition de phase) dans le cas où la majorité de l'énergie était effectivement contenue en basses fréquences. Le résultat de Edmonds et Culling (2009) pourrait ainsi expliquer la sous-estimation des modèles (hypothèse appuyée par Miranda et Cabrera (2008)) et constitue de ce fait un point important à approfondir. En effet, dès lors que l'on s'intéresse à des sources réelles ou que la diffusion acoustique est prise en compte, la corrélation interaurale est un indice acoustique fortement affecté. Il serait donc intéressant de mener une série d'études pour déterminer plus précisément l'effet de la corrélation interaurale sur la sonie dans des conditions réalistes d'écoute et d'apporter les modifications nécessaires aux modèles afin de prendre en compte cet effet.

Puisque les différences entre les fonctions de sonie monaurales individuelles n'expliquent pas les différentes stratégies de sommation binaurales, d'autres facteurs (probablement d'ordre cognitif) doivent être envisagés. Une hypothèse qui semble intéressante à considérer est celle de la constance perceptuelle, qui par analogie avec l'invariance du jugement de sonie avec la distance rapportée par Zahorik et Wightman (2001), impliquerait l'invariance du jugement de sonie avec l'angle d'incidence de la source. Les DLS mesurées ne sont pas clairement représentatives de ce phénomène (qui se traduirait par des DLS toujours nulles), mais il est possible que ce phénomène soit mis en jeu chez certains auditeurs. La mise en évidence de l'effet de constance perceptuelle en sonie nécessiterait d'utiliser plusieurs sources de différentes puissances et de guider l'auditeur pour qu'il porte son jugement sur la puissance de la source (consigne "objective"). Le recours à des sources réelles (non-stationnaires), des indices de localisation visuels ou encore des indices dynamiques (si la position de la source varie) devrait favoriser l'apparition d'un tel phénomène.

Un troisième point plus général qui nous semble important concerne la relation entre l'analyse de scène auditive (organisation de l'information par groupement ou ségrégation, en objets sonores) et la sonie. Par exemple, dans le chapitre 4, l'effet de la distribution spatiale de deux sources tonales sur la sonie a été étudié. Cependant, le fait que la séparation spatiale des deux sources ait entraîné (ou non) la ségrégation des deux sources d'un point de vue perceptif n'est pas clair. Il aurait été intéressant de demander aux auditeurs de rapporter leur percept ou d'utiliser des modulations d'amplitude indépendantes pour chaque source, n'influant pas sur la sonie mais permettant de forcer le groupement ou la ségrégation des sources. L'incertitude sur ce point est en tous cas une limite à l'interprétation du résultat en sonie. D'autre part, la dernière partie de l'expérience, dans laquelle les auditeurs devaient comparer la sonie globale de la source bitonale (1 kHz + 5 kHz) avec la sonie d'une seule composante tonale (1 kHz ou 5 kHz) est un parfait exemple de l'influence des mécanismes de l'analyse de scène auditive sur la sonie. En effet, comme cela rendait la tâche plus facile, les auditeurs avaient a priori tendance à grouper la composante seule avec la même composante dans le son complexe, faussant ainsi le jugement de sonie globale. Ces observations rejoignent les travaux de Grimault *et al.* (2007) et McAdams *et al.* (1998), mettant en avant l'importance de la prise en compte de ces mécanismes pour prédire la sonie dans des champs acoustiques complexes. Ce sujet est vaste et peu exploré, ouvrant ainsi des perspectives à long terme.

Audiogrammes

Une partie des expériences présentées dans la thèse ayant été réalisée à Darmstadt en Allemagne (Chapitre 5), la liste des audiogrammes est présentée en deux parties.

A.1 Laboratoire Vibrations Acoustique, INSA-Lyon

Les audiogrammes ont été réalisés à l'aide d'un dispositif Oscilla USB-300 composé d'un casque SILENTA dédié (avec audiomètre intégré), d'un bouton de réponse et du logiciel AudioConsole version 2.3.8. Le test utilisé permettait de tester rapidement l'audition des auditeurs à 20 dB HL, seuil considéré comme limite admise pour un normo-entendant. Le niveau de départ pour chaque fréquence correspondait donc à 20 dB HL et le niveau maximum utilisé était fixé à 80 dB HL. Les stimuli utilisés étaient des sons purs aux fréquences 125, 250, 500, 750, 1000, 1500, 2000, 3000, 4000, 6000, 8000 Hz de durée 1 s. L'intervalle entre deux stimuli variait de manière aléatoire dans l'intervalle [1 s - 4 s]. Le test commençait par les mesures à 1 kHz à l'oreille gauche puis 1 kHz à l'oreille droite. Pour toutes les autres mesures, la fréquence et le canal étaient ensuite choisis de manière aléatoire.

Tant que le son n'était pas détecté par l'auditeur, le niveau de celui-ci augmentait par pas de 5 dB. Une fois le stimulus détecté, le niveau HL courant était enregistré comme seuil audiométrique. Si aucun son n'était détecté, le seuil audiométrique retenu était la limite haute (80 dB HL).

Le détail des résultats obtenus pour l'ensemble des auditeurs ayant participé aux expériences à Lyon est présenté dans le tableau A.1. Seuls les résultats pour les fréquences comprises entre 500 Hz et 6 kHz sont recensés ici sachant qu'aucune perte n'a été relevée aux fréquences exclues du tableau.

Auditeurs	Age	Sexe	Oreille	Fréquence							
				500 Hz	750 Hz	1 kHz	1.5 kHz	2 kHz	3 kHz	4 kHz	6 kHz
MV	25	M	G	-	-	-	-	-	-	-	-
			D	-	-	-	-	-	-	-	-
EP	49	M	G	-	-	-	-	-	-	-	-
			D	-	-	-	-	-	-	-	-
MP	34	F	G	-	-	-	-	-	-	-	25
			D	-	-	-	-	-	-	-	-
TL	25	M	G	-	-	-	-	-	-	-	-
			D	-	-	-	-	-	-	-	-
CM	24	F	G	-	-	-	-	-	-	-	-
			D	-	-	-	-	-	-	-	-
LB	33	M	G	-	-	-	-	-	-	-	-
			D	-	-	-	-	-	-	-	-
LC	28	F	G	-	-	-	-	-	-	-	-
			D	-	-	-	-	-	-	-	-
CS	25	F	G	-	-	-	-	-	-	-	-
			D	-	-	25	-	-	-	-	-
RS	25	M	G	-	-	30	-	-	-	-	-
			D	-	-	-	-	-	-	-	-
CB	25	M	G	-	-	-	30	25	25	-	-
			D	25	25	-	30	35	-	25	-
SB	26	F	G	-	-	-	-	-	-	-	-
			D	-	-	-	-	-	-	-	40
WA	26	M	G	-	-	-	-	-	-	-	-
			D	-	-	-	-	-	-	30	25
SL	24	F	G	-	-	-	-	-	-	-	-
			D	-	-	-	-	-	-	-	25
LG	24	M	G	-	-	-	-	-	-	-	-
			D	-	-	-	-	-	-	-	25
FA	27	M	G	-	-	-	-	-	-	-	25
			D	-	-	-	-	-	-	-	-
MY	27	F	G	-	-	25	-	25	-	-	-
			D	-	-	-	-	-	-	-	-
ML	26	F	G	-	-	-	-	-	-	-	-
			D	-	-	-	-	-	-	-	-
JM	27	F	G	-	-	-	-	-	-	-	-
			D	-	-	-	-	-	-	-	-

Tableau A.1: Seuils audiométriques (dB HL) de tous les auditeurs ayant participé aux expériences (à Lyon). L'absence de valeur désigne un seuil inférieur ou égal à 20 dB HL.

A.2 Institut für psychologie, TU-Darmstadt

Les audiogrammes ont été réalisés à l'aide d'un système MADSEN Astera contrôlé par le logiciel OTOsuite. Les sons étaient restitués à l'aide d'un casque Sennheiser HDA 200. Les tests audiométriques ont été réalisés manuellement par octaves (aux fréquences 125, 250, 500, 1000, 2000, 3000, 4000 et 8000 Hz), et pour chaque oreille. Les stimuli étaient des sons purs ininterrompus. La procédure était la suivante : une fréquence était choisie "au hasard", le niveau de départ était fixé à 0 dB HL, puis le niveau était augmenté par pas de 5 dB par intervalles de temps aléatoires jusqu'à ce que le son soit détecté par l'auditeur. Ce dernier niveau était alors considéré comme le seuil audiométrique pour la fréquence sélectionnée. Le test était d'abord mené pour l'oreille gauche, puis pour l'oreille droite.

Le détail des résultats obtenus pour tous les auditeurs ayant participé aux expériences menées à Darmstadt est présenté dans le tableau [A.2](#).

Auditeurs	Age	Sexe	Oreille	Fréquence							
				125 Hz	250 Hz	500 Hz	1 kHz	2 kHz	3 kHz	4 kHz	8 kHz
WE	57	M	G	-	-	-	25	-	-	-	-
			D	-	-	-	-	-	-	-	-
JS	28	M	G	-	-	-	-	-	-	-	-
			D	-	-	-	-	-	-	-	-
FL	25	M	G	-	-	-	-	-	-	-	-
			D	-	-	-	-	-	-	-	-
LB	24	M	G	-	-	-	-	-	-	-	-
			D	-	-	-	-	-	-	-	-
JR	20	M	G	-	-	-	-	-	-	-	-
			D	-	-	-	-	-	-	-	-
JL	24	M	G	-	-	-	-	-	-	-	-
			D	-	-	-	-	-	-	-	-
TH	26	F	G	-	-	-	-	-	-	-	-
			D	-	-	-	-	-	-	-	-
NM	19	F	G	-	-	-	-	-	-	-	-
			D	-	-	-	-	-	-	-	-
MG	32	M	G	-	-	-	-	-	-	-	-
			D	-	-	-	-	-	-	-	-
JB	25	F	G	-	-	-	-	-	-	-	-
			D	-	-	-	-	-	-	-	-
AK	19	F	G	-	-	-	-	-	-	-	-
			D	-	-	-	-	-	-	-	-
ED	20	F	G	-	-	-	-	-	-	-	-
			D	-	-	-	-	-	-	-	-
TT	20	M	G	-	-	-	-	-	-	-	-
			D	-	-	-	-	-	-	-	-
TG	21	M	G	-	-	-	-	-	-	-	-
			D	-	-	-	-	-	-	-	-
LW	21	F	G	-	-	-	-	-	-	-	-
			D	-	-	-	-	-	-	-	-
AG	28	F	G	-	-	-	-	-	-	-	-
			D	-	-	-	-	-	-	-	-
MV	27	M	G	-	-	-	-	-	-	-	-
			D	-	-	-	-	-	-	-	-

Tableau A.2: Seuils audiométriques (dB HL) de tous les auditeurs ayant participé aux expériences (à Darmstadt). L'absence de valeur désigne un seuil inférieur ou égal à 20 dB HL.

Synthèse binaurale générique

B.1 Principe général

Les techniques binaurales permettent de simuler (par reproduction ou synthèse) des sources acoustiques dans un environnement réel au casque, en restituant aux oreilles de l'auditeur une copie la plus fidèle possible des signaux acoustiques qui auraient été engendrés dans une situation réelle d'écoute. Deux principaux types de techniques binaurales peuvent être discernés : la restitution d'enregistrements binauraux de scènes sonores, et la synthèse binaurale consistant à convoluer un son monophonique par une paire de réponses impulsionnelles binaurales (G/D) mesurées en champ libre (ou dans une acoustique particulière que l'on souhaite reproduire).

Dans le cadre des mesures de sonie impliquées dans les chapitres 2 et 4, l'objectif était de simuler des sources de bruit artificielles dans un environnement réel, à 2 m de distance de la tête de l'auditeur et dans 13 directions différentes de l'espace. Or, il a été montré par Sivonen *et al.* (2005) et Sivonen et Ellermeier (2008), que les égalisations en sonie obtenues en utilisant la synthèse binaurale étaient similaires à celles obtenues en champ réel. Nous avons donc fait le choix d'utiliser la synthèse binaurale afin de pouvoir nous affranchir de l'utilisation permanente d'une structure encombrante (non envisageable dans le cadre du laboratoire) d'une part et de mieux maîtriser les signaux restitués aux oreilles des auditeurs d'autre part.

Lors de leur propagation depuis la source jusqu'aux oreilles, les sons incidents sont absorbés et diffusés par le corps de l'auditeur (buste, tête, pavillons,...). Ce filtrage acoustique crée des différences spectrales, de niveau et de temps d'arrivée entre les signaux reçus à chaque oreille. Ces différences interaurales sont caractéristiques de la position de la source dans le cas du champ libre. L'ensemble de cette information peut être contenue dans une série de filtres connus sous le nom de Head-Related Transfer Functions (HRTFs). Pour la synthèse binaurale, il est alors possible d'utiliser soit des HRTFs individuelles mesurées sur les différents auditeurs, soit des HRTFs dites "génériques" mesurées sur une tête artificielle. L'idée initiale était de réaliser les deux types de synthèse pour pouvoir évaluer la variabilité interindividuelle dans les DLS, induite par les différences d'HRTFs individuelles. C'est la raison pour laquelle nous avons entrepris de construire une structure avec plusieurs haut-parleurs pour les différentes élévations et que les mesures ont été réalisées dans une salle semi-anéchoïque, la seule salle à notre disposition permettant d'accueillir la structure (voir § B.2). Cependant, le dispositif que nous avons commandé pour réaliser les mesures individuelles à l'entrée du conduit auditif (microphones à électret intégrés dans des bouchons d'oreille en silicone) n'était pas conforme à notre demande et n'a pas permis de réaliser des mesures suffisamment précises.

Seules les mesures réalisées sur la tête artificielle ont donc finalement pu être exploitées. L'avantage de la synthèse générique réalisée dans le cadre de la thèse est qu'elle permet de générer ap-

proximativement les mêmes différences interaurales pour tous les auditeurs et donc évite de devoir prendre en compte les différences d'HRTFs individuelles dans la comparaison des résultats. Cela permet également de s'affranchir d'une source d'erreur expérimentale supplémentaire, susceptible d'accroître la variabilité interindividuelle de manière indésirable dans les mesures de sonie. De plus, la finalité de l'étude était de proposer un modèle permettant de prédire la sonie binaurale à partir de mesures acoustiques réalisées sur un mannequin acoustique. Le recours à la synthèse binaurale générique était donc justifié.

La construction des paires de filtres (G/D) à convoluer aux stimuli pour chaque direction à simuler, se décompose en deux principales étapes : la mesure des réponses impulsionnelles binaurales (RIB) sur la tête artificielle pour chacune des directions et l'égalisation des différents transducteurs utilisés dans la chaîne (de la mesure des RIBs à la restitution au casque). Le principe général du processus de synthèse binaurale mis en place est présenté dans la figure B.1, les différentes étapes sont ensuite détaillées dans les parties B.2, B.3 et B.4.

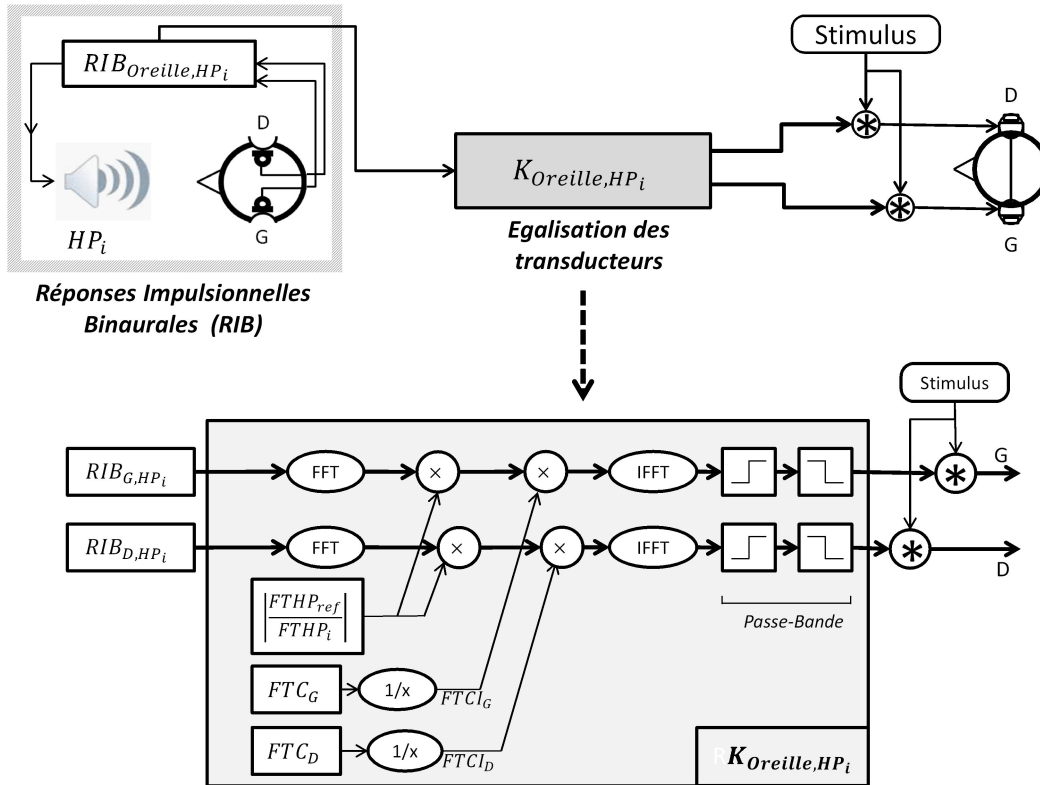


FIGURE B.1: Principe général du processus de synthèse binaurale. Le cadre du bas détaille les opérations d'égalisation effectuées sur les RIBs pour construire les paires de filtres à convoluer aux stimuli pour la synthèse.

Il convient de compenser la réponse et la sensibilité des microphones utilisés, la réponse du casque couplé avec la tête (fonction de transfert entre le signal électrique délivré au casque et la pression "aux tympans"), ainsi que la réponse des haut-parleurs utilisés pour les mesures de RIB. Dans le cas des microphones, il a été suggéré par Møller (1992) de réaliser les mesures de RIB et

les mesures de fonction de transfert du casque avec les mêmes microphones. Ainsi, au cours de la déconvolution des RIBs par la réponse du casque (afin de s'affranchir de celle-ci), les fonctions de transfert des microphones se compensent. Dans le cas présent toutes les mesures ont été réalisées avec les microphones intégrés de la TA, l'égalisation des microphones a donc été réalisée par défaut.

B.2 Mesure des réponses impulsionnelles binaurales (RIB)

B.2.1 Dispositif expérimental

Les mesures des RIBs ont été effectuées dans une salle semi-anéchoïque du CETIAT¹ (sol non traité).

Cinq haut-parleurs (Visaton DSM 50-FFL-8) ont été fixés sur une structure suivant un arc de cercle de rayon 197 cm \pm 5 mm, centré sur la tête du mannequin acoustique (cf. figure B.2, photo à gauche). Cette distance permettait de se placer dans les conditions de l'approximation d'ondes planes, nécessaire pour que les HRTFs ne dépendent plus de la distance de la source. La structure métallique a été entièrement recouverte de mousse acoustique absorbante afin de limiter les phénomènes de réflexion. Du point de vue du mannequin, les angles d'élévation décrits par les différents haut-parleurs par rapport au plan horizontal étaient les suivants : 0° (en face de l'auditeur), 30°, 45°, 60° et 90° (au zénith).

Un siège situé au centre du dispositif autorisait un réglage de la hauteur de l'assise afin de faire correspondre le centre de la tête artificielle avec le centre du quart de cercle décrit par les haut-parleurs. Un système de blocage situé sur l'axe de rotation du siège permettait une rotation de 360° par pas de 30° dans le plan horizontal (cf. figure B.2, photo au centre droit). Seuls six angles d'azimut compris entre 0° et 180° (source à main gauche) ont été retenus pour les mesures (0°, 30°, 60°, 90°, 150°, 180°), en admettant la symétrie des résultats par rapport au plan sagittal.

La synthèse des stimuli ainsi que l'acquisition des signaux était assurée par le logiciel ARTA² configuré pour fonctionner avec une carte son Digigram VxPocket 440 (44100 Hz, 24 bits). Une fois générés, les signaux étaient amplifiés via un amplificateur QD-4240 puis aiguillés vers les différents haut-parleurs par l'intermédiaire d'un sélecteur d'écoute Sima SSW-L6 EX. L'enregistrement binaural était ensuite réalisé à l'aide d'une tête artificielle Cortex MK1 équipée de deux microphones 1/2" intégrés dans les canaux auditifs.

B.2.2 Mesures

En accord avec Mateljan et Ugrinović (2003) et Stan *et al.* (2002), les mesures ont été réalisées à l'aide d'un sinus glissant logarithmique, signal excitateur le plus efficace pour la mesure de réponses impulsionnelles dans un environnement non bruyant et invariant dans le temps. Ses propriétés ont également l'avantage de permettre la séparation de la partie linéaire et de la partie non-linéaire (distorsions harmoniques) de la réponse du système (Farina, 2000). Le sinus glissant était défini sur 256 kbits correspondant à une durée totale de 5.8 s à 44,1 kHz.

1. CETIAT : Centre Technique des Industries Aéronautiques et Thermiques, 25, avenue des Arts - BP 52042 - 69603 VILLEURBANNE cedex

2. ARTA Software : Audio Measurement and Analysis Software from ARTALABS - J. Rodina 4, 21215 Kastel Luksic, Croatia. Auteur : Mateljan, I.



FIGURE B.2: Parties du dispositif expérimental mis en place pour les mesures des réponses impulsionnelles binaurales.

Chaque mesure consistait à enregistrer le signal d'un microphone de la TA (G ou D) pour une direction donnée en réponse à l'excitation délivrée par un des haut-parleurs, chaque direction étant caractérisée par un angle d'azimut $[0^\circ, 30^\circ, 60^\circ, 90^\circ, 150^\circ, 180^\circ]$ et un angle d'élévation $[0^\circ, 30^\circ, 45^\circ, 60^\circ, 90^\circ]$. Au total, 30 paires (G/D) de réponses impulsionnelles binaurales ont été réalisées sur la TA (6 azimuts x 5 élévations). Chaque RIB calculée représentait alors la réponse de l'ensemble de la chaîne audio du dispositif expérimental (chaîne d'amplification, transduction des HPs, propagation du son dans l'air, filtrage de l'auditeur (HRTF), transduction des microphones, chaîne d'amplification).

Un fenêtrage temporel a ensuite été opéré sur chaque RIB afin d'extraire la partie de la réponse impulsionnelle contenant la plus grande partie possible de l'information permettant de caractériser notre système, tout en essayant de s'affranchir de l'énergie apportée par les éventuelles réflexions acoustiques. Les mesures ont été réalisées en chambre semi-anéchoïque, induisant l'apparition d'une source image due à la réflexion sur le sol. Le temps d'arrivée de cette source image par rapport au son direct était donc prépondérant sur le choix de la taille de la fenêtre. Ce temps d'arrivée étant minimal pour le HP1 situé dans le plan horizontal, la taille de la fenêtre était toujours inférieure au temps d'arrivée de la réflexion mesuré pour cette condition. Quelques réflexions de moindre amplitude étaient également observées, probablement entraînées par les différents dispositifs (chaise, socle du

mannequin, structure,...etc). La taille de la fenêtre (de Tuckey) variait ainsi entre 1.8 et 2.3 ms, permettant d'exclure avec certitude la première réflexion du sol ainsi que les autres réflexions les plus gênantes.

Des jeux d'HRTFs ont été dérivés des RIBs pour chaque oreille, en divisant leurs spectres par la fonction de transfert de l'ensemble du système de mesure, mesurée à l'aide d'un microphone PCB 1/4" au centre de la tête en son absence (Møller, 1992). Les HRTFs, caractéristiques acoustiques propres à la tête artificielle, ont permis de comparer nos mesures aux HRTFs contenues dans d'autres bases de données. L'allure et l'ordre de grandeur des HRTFs mesurées pour la tête Cortex MK1 étaient par exemple en accord avec la base de données du MIT (Gardner *et al.*, 1994) établie pour une tête artificielle KEMAR, Knowles Electronics DB-4004. Nos HRTFs étaient également proches des HRTFs rapportées par Sivonen et Ellermeier (2008) dans le plan horizontal pour le mannequin B&K Type 4100, bien que le maximum d'amplitude correspondant à la résonance du conduit auditif (autour de 2.5 kHz pour les TAs Cortex Mk1 et KEMAR) se situait plus haut en fréquence pour la tête B&K (entre 4 et 5 kHz).

B.3 Égalisation des transducteurs

B.3.1 Fonctions de transfert du casque (FTCs)

Afin de s'affranchir de la fonction de transfert du casque et du couplage du casque avec la tête de l'auditeur lors de la restitution, il est nécessaire de convoluer les RIBs avec la fonction de transfert du casque inverse (FTCI) pour l'oreille correspondante.

Les mesures des FTCs ont été réalisées sur la TA pour le casque Sennheiser HD650 neuf utilisé par la suite pour les mesures de sonie. Le dispositif et la méthode de mesure étaient strictement identiques à ceux utilisés pour les mesures des RIBs, à la seule différence que les stimuli étaient cette fois-ci diffusés au casque (canal G ou D) positionné et ajusté sur la tête du mannequin. Chaque mesure de réponse impulsionnelle était répétée trois fois, avec un repositionnement du casque entre chaque répétition. Comme pour les RIBs, un fenêtrage a été opéré pour chaque RI à l'aide d'une fenêtre de Tuckey de 2.3 ms, incluant une croissance et une décroissance de 0.2 ms. Les FTCs obtenues pour les trois répétitions de la mesure sont présentées dans la figure B.3 pour chaque oreille (G et D).

La variation des FTCs sur les différents repositionnements du casque est quasiment nulle sur l'ensemble du spectre jusqu'à 10 kHz pour l'oreille D. Des variations faibles (< 2 dB) sont cependant observées pour l'oreille G autour de 8 kHz.

Une FTC moyenne a ensuite été déterminée pour chaque oreille, en moyennant les spectres calculés pour chaque répétition. Finalement, la correction apportée aux RIBs (dans le domaine fréquentiel) correspondait à la fonction de transfert inverse (FTCI) :

$$CorrectionCasque(Oreille) = FTCl_{G,D} = \frac{1}{FTC_{G,D}}. \quad (B.1)$$

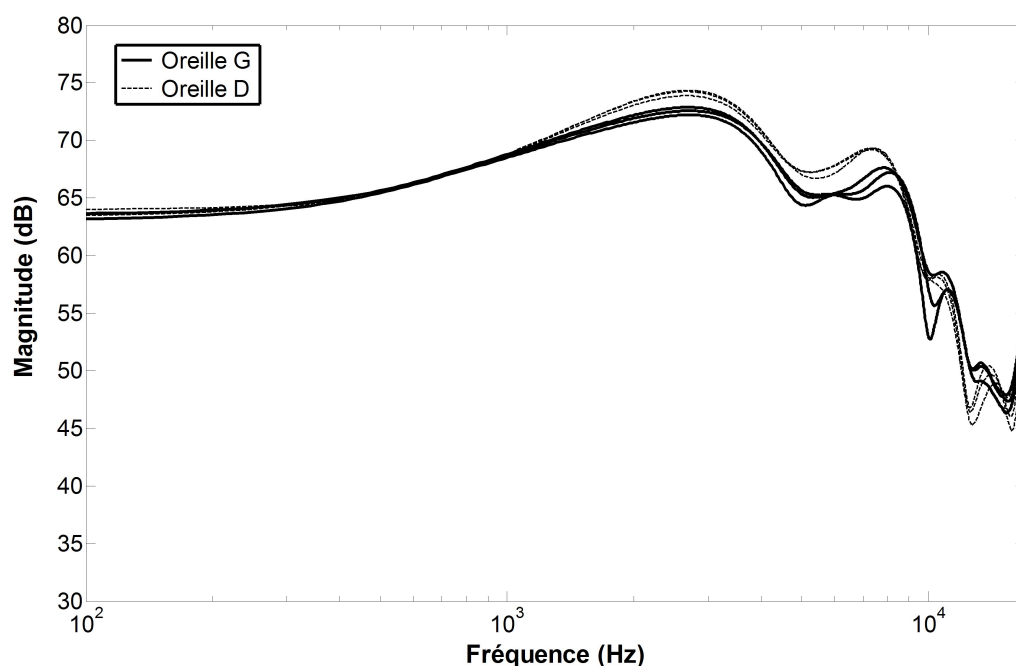


FIGURE B.3: Fonctions de transfert du casque calculées pour 3 repositionnements sur la TA, pour l'oreille G (courbes continues) et pour l'oreille D (courbes en pointillé).

B.3.2 Fonctions de transfert des haut-parleurs (FTHPs)

Les mesures de RIB ont été effectuées à l'aide de 5 haut-parleurs différents non-appairés correspondant aux 5 angles d'élévation ($HP_1 : [\Theta = 0^\circ]$, $HP_2 : [\Theta = 30^\circ]$, $HP_3 : [\Theta = 45^\circ]$, $HP_4 : [\Theta = 60^\circ]$, $HP_5 : [\Theta = 90^\circ]$). Il a donc été nécessaire de mesurer les fonctions de transfert de ces différents haut-parleurs (FTHPs) afin de s'affranchir des différences de filtrage induites entre les différentes élévations dans le processus de synthèse.

Une réponse impulsionnelle a été mesurée pour chaque haut-parleur dans une chambre semi-anéchoïque du laboratoire. Le même système de mesure et de calcul des RIs que celui utilisé lors des mesures de RIBs ou des FTCs (cf. partie B.2) a été mis en place. L'enregistrement était réalisé à l'aide d'un microphone PCB placé à 1.97m +/- 1cm en face du haut-parleur à analyser, à 1,20 m du sol. Des bandes de mousse acoustique absorbante étaient disposées au sol entre le micro et le HP afin d'atténuer la source image causée par la réflexion au sol.

Pour chaque HP, la mesure était répétée trois fois. Chaque RI calculée était fenêtrée par une fenêtre de Tuckey de 2.3 ms puis transposée dans le domaine de Fourier. La fonction de transfert finale pour chaque HP était obtenue par moyennage des trois fonctions de transfert correspondantes. Les fonctions de transfert obtenues pour chacun des cinq haut-parleurs sont représentées dans la figure B.4.

Afin d'harmoniser les différences entre les spectres d'amplitudes mises en évidence dans la figure B.4, des corrections relatives (par rapport au HP_1 d'élévation 0°) ont été appliquées aux RIBs en fonction du HP concerné. Ainsi, une réponse à phase minimale a été extraite de chaque FTHP, en accord avec la procédure proposée par Hawksford (1997). La correction à appliquer aux RIBs correspondait alors au rapport des FTHPs à phase minimale du HP_1 de référence et du HP_i utilisé

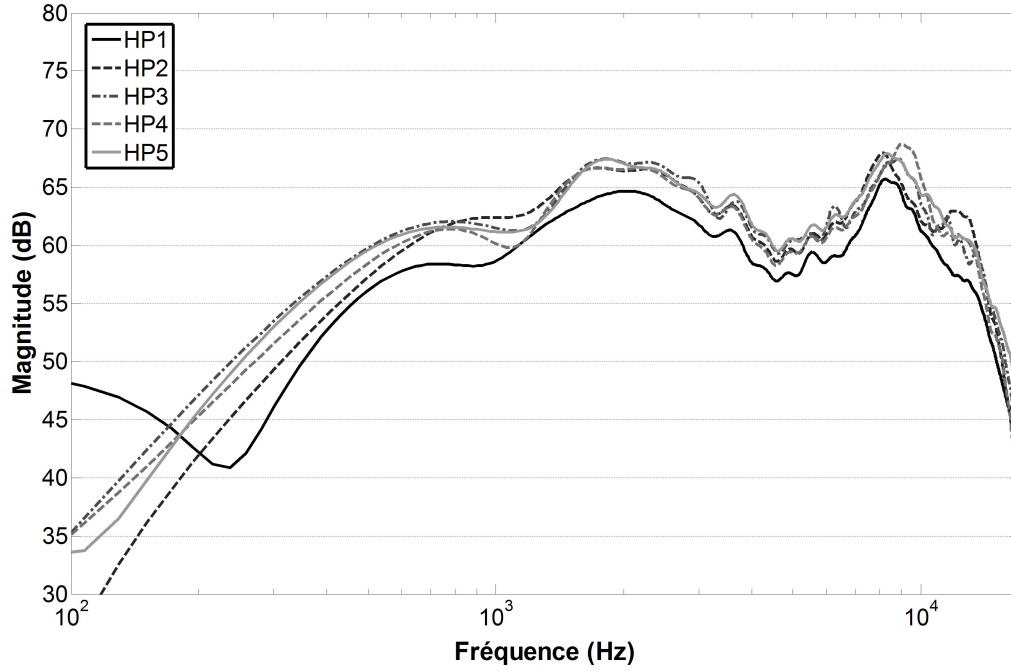


FIGURE B.4: Fonctions de transfert obtenues pour les cinq haut-parleurs (FTHPs). Les HPs 1, 2, 3, 4 et 5 désignent respectivement les HPs d'élévation 0°, 30°, 45°, 60° et 90°.

pour mesurer la RIB :

$$CorrectionHP(HP_i) = \frac{FTHP_{i_{phaseminimale}}}{FTHP_{i_{phaseminimale}}} \quad (B.2)$$

B.4 Construction des filtres pour la synthèse

Au final, 30 paires de filtres (G, D) ont été conçues pour permettre la synthèse binaurale dans 30 directions différentes. La construction des filtres consistait simplement à corriger les RIBs initialement mesurées afin d'égaliser la réponse des différents transducteurs. Un filtrage passe-bande entre 300 Hz et 6 kHz était aussi systématiquement appliqué à la fin du processus de construction des filtres, afin d'atténuer les bandes de fréquences en dehors du domaine d'intérêt pour notre étude.

Les fréquences de coupure du passe-bande ont été déterminées principalement par la fréquence de coupure basse estimée des HPs ($\simeq 300$ Hz) d'une part, et la fréquence haute à partir de laquelle les HRTFs commençaient à se dégrader pour l'oreille contralatérale et pour laquelle la symétrie des HRTFs (G, D) n'était plus respectée.

L'expression générale du calcul des filtres pour la synthèse FS est présentée dans l'équation B.3.

$$FS_{\Phi, \Theta, Oreille} = ifft(fft(RIB_{\Phi, \Theta, Oreille}) \times CorrectionHP_{\Theta} \times CorrectionCasque_{Oreille} \times fft(PasseBande)) \quad (B.3)$$

A titre d'exemple, les spectres d'amplitude des paires de filtres FS obtenues dans le plan horizontal sont représentés dans la figure B.5.

Pour réaliser la synthèse binaurale d'une source dans une direction donnée, il suffisait alors de convoluer le stimulus monaural souhaité avec chacun des filtres (FS) G et D correspondant à cette direction.

B.5 Validation de la synthèse

Afin de valider la synthèse, les signaux enregistrés "aux tympans" sur la tête artificielle pour une source frontale dans un environnement réel ont été comparés aux signaux enregistrés pour la même source restituée au casque suite au processus de synthèse binaurale. La comparaison a été effectuée pour différents stimuli (bruit rose, voix, ...etc). Les résultats étant très similaires, seul le cas du bruit rose sera présenté ici.

B.5.1 Enregistrement en champ réel

Dans un premier temps, des enregistrements ont été effectués en champ réel dans une chambre semi-anéchoïque du laboratoire.

Les bruits roses étaient diffusés sur le HP_1 (HP utilisé pour l'élévation 0° dans les mesures de RIBs) placé à 1.97 m du centre de la TA en incidence frontale, puis enregistrés "aux tympans" par les microphones de la TA. Un filtrage passe-bande identique à celui appliqué dans le processus de synthèse (cf. partie B.4) a ensuite été appliqué sur chaque canal (G et D) afin de pouvoir comparer ces enregistrements à ceux mesurés lors de la restitution au casque après la synthèse (cf. section B.5.2).

B.5.2 Enregistrement à la restitution (post-synthèse)

Une seconde partie a ensuite consisté à enregistrer les signaux G et D toujours "aux tympans" sur la TA, mais cette fois-ci, lors de la restitution au casque de la même source frontale (bruit rose) à la suite du processus de synthèse.

Il a été nécessaire au préalable, de mesurer les RIBs dans la chambre semi-anéchoïque utilisée pour les enregistrements en champ réel. Ainsi, les RIBs G et D pour l'incidence frontale ($\Phi = 0^\circ$, $\Theta = 0^\circ$) ont été mesurées sur la TA à l'aide du même dispositif (strictement) que pour les mesures de RIBs initiales (cf. section B.2). Le bruit rose a ensuite été convolué à la paire de filtres pour la synthèse (FS), calculés conformément au processus de synthèse décrit dans la section B.4. Les signaux ainsi synthétisés ont finalement été restitués au casque à la TA et enregistrés "aux tympans".

B.5.3 Comparaison des enregistrements

Une comparaison des deux enregistrements (en situation d'écoute réelle et en situation d'écoute virtuelle) a ensuite été réalisée dans le but d'estimer la fidélité de la synthèse (pour l'oreille G, puis pour l'oreille D). Préalablement recalés temporellement par corrélation croisée, les deux enregistrements ont ensuite été comparés dans le domaine fréquentiel (spectres de puissance RMS) par

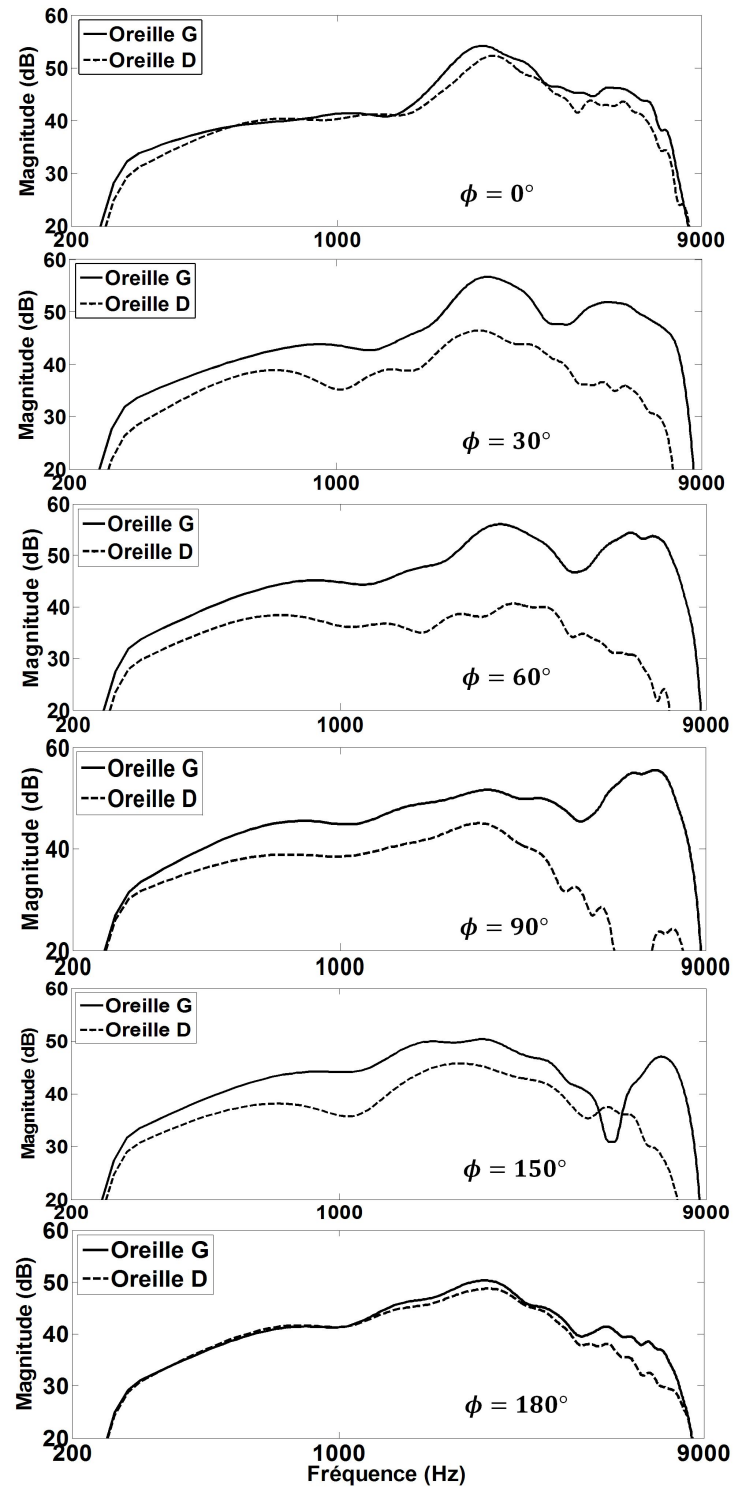


FIGURE B.5: Spectres d'amplitude des paires de filtres *FS* calculés pour la synthèse, pour chaque azimut, dans le plan horizontal.

bandes de 1/3 d'octave. Étant donné que les niveaux moyens de diffusion et d'enregistrement des deux chaînes de mesure n'ont pas été calibrés entre eux, une égalisation des énergies moyennes a été réalisée sur les spectres de puissance à comparer. En tenant compte de la bande passante du filtre appliqué dans le processus de synthèse, comprise entre 300 Hz et 6 kHz, seules les bandes de fréquences comprises entre le tiers d'octave centré à 400 Hz et le tiers d'octave centré à 5 kHz (inclus), ont été prises en compte dans le calcul des énergies moyennes. La figure B.6 présente les erreurs de niveau entre les deux enregistrements par bande de tiers d'octave (après l'égalisation des niveaux moyens).

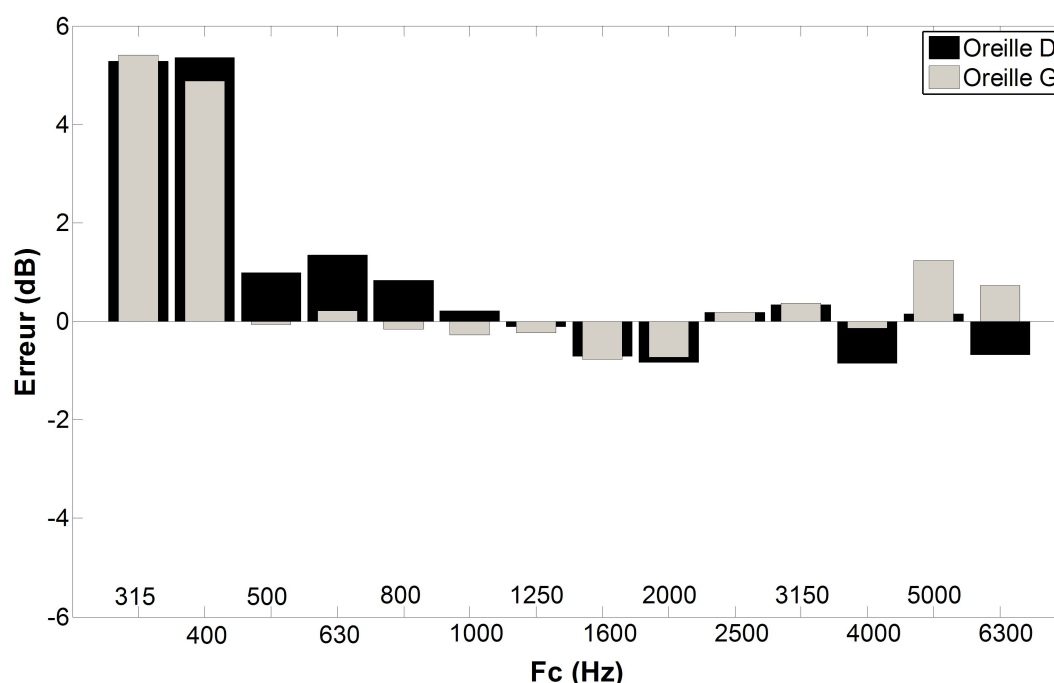


FIGURE B.6: Erreurs de niveau par bande de tiers d'octave entre les enregistrements "aux tympans", en situation d'écoute réelle et en situation d'écoute virtuelle (synthèse binaurale). Les barres grises et noires représentent respectivement les résultats pour l'oreille G et pour l'oreille D.

Pour les bandes de fréquences centrées entre 500 Hz et 6300 Hz, toutes les erreurs étaient strictement inférieures à 2 dB. L'erreur maximale était de 1.34 dB pour la bande centrée à 630 Hz à l'oreille D. En revanche, pour les bandes de tiers d'octave centrées à 315 Hz et à 400 Hz, les erreurs étaient plus grandes, atteignant 5 dB. Une différence entre les deux méthodes était l'ordre d'application du filtre passe-bande. En effet, dans le cas du "champ réel", le passe-bande était appliqué après l'enregistrement alors que pour la synthèse, le filtrage était appliqué avant l'enregistrement. Cependant, il nous a semblé que si cela devait être source d'erreur, ces erreurs auraient été plus susceptibles d'apparaître en-dessous de 300 Hz (fréquence de coupure basse du passe-bande). Aucune explication claire n'a en tous cas été trouvée.

Les erreurs observées entre 500 Hz et 6300 Hz correspondaient en tous cas à l'ordre de grandeur des erreurs rapportées par Sivonen *et al.* (2005) pour la synthèse binaurale individuelle. La méthode de vérification mise en place dans Sivonen *et al.* (2005) était une peu différente : synthèse d'un bruit blanc, déconvolution du bruit blanc d'origine, puis comparaison du résultat avec la RIB originale.

Mais les erreurs calculées étaient comprises d'une manière générale dans un intervalle de ± 2 dB pour l'ensemble des auditeurs.

Nous pouvons donc conclure que la synthèse binaurale permet de simuler précisément les niveaux de pression "aux tympans" qui auraient été produits par une source réelle en incidence frontale. Cette conclusion a été établie pour les bandes de tiers d'octave centrées entre 500 Hz et 6300 Hz. Étant donné que les protocoles expérimentaux et le processus de synthèse étaient identiques pour l'ensemble des angles d'incidence, il semble raisonnable de généraliser ce résultat, en admettant que la synthèse binaurale est valide quelle que soit la direction.

Calibration du casque

La calibration du casque consiste à déterminer le niveau du signal électrique à fournir au casque, pour produire le niveau de pression acoustique souhaité aux oreilles de l'auditeur. Si le signal considéré est la pression acoustique mesurée aux tympans (membrane du microphone positionnée au fond du conduit auditif sur la tête artificielle), on parlera alors de calibration aux tympans. Bien que ce ne soit pas exactement le cas sur la tête artificielle Cortex MK1 (microphones situés à une position intermédiaire du conduit), nous utiliserons tout de même l'expression "aux tympans" pour désigner cette position, par approximation.

Dans le cadre des mesures de sonie de sources spatialisées, le but était de reproduire aux tympans de l'auditeur les niveaux de pressions auxquels il aurait été soumis en présence d'une source réelle, émettant à un certain niveau acoustique fixe. Si la source est considérée dans les conditions de champ libre, face à la tête de l'auditeur (incidence frontale), on parlera de calibration en "équivalent champ libre". Le niveau de pression de référence de la source est alors mesuré au centre de la tête artificielle en son absence. Mais si la source de référence est considérée dans les conditions de champ diffus (ce qui était le cas pour la calibration de la plupart des expériences présentées), on parlera de calibration en "équivalent champ diffus". Le niveau de pression de référence de la source peut alors être mesuré en tout point de l'espace.

C.1 Calibration "aux tympans", en équivalent champ diffus

Les principales étapes de calibration du casque "aux tympans" en équivalent champ diffus sont détaillées dans cette partie. La description du protocole est basée sur l'exemple des sources artificielles utilisées dans les chapitres 2 et 4, synthétisées à l'aide d'un bruit rose filtré par bandes de tiers d'octave centrées à 1 kHz, 2236 Hz ou 5 kHz.

Le protocole de calibration était décomposé en deux principales étapes : la mesure des niveaux de pression équivalents "aux tympans" en champ diffus, puis la calibration du casque à proprement parler.

C.1.1 Mesure des niveaux de pression équivalents "aux tympans"

Le schéma présenté dans la figure C.1 représente le dispositif mis en place dans la salle réverbérante du laboratoire, afin de mesurer les niveaux de pression équivalents "aux tympans", en champ diffus.

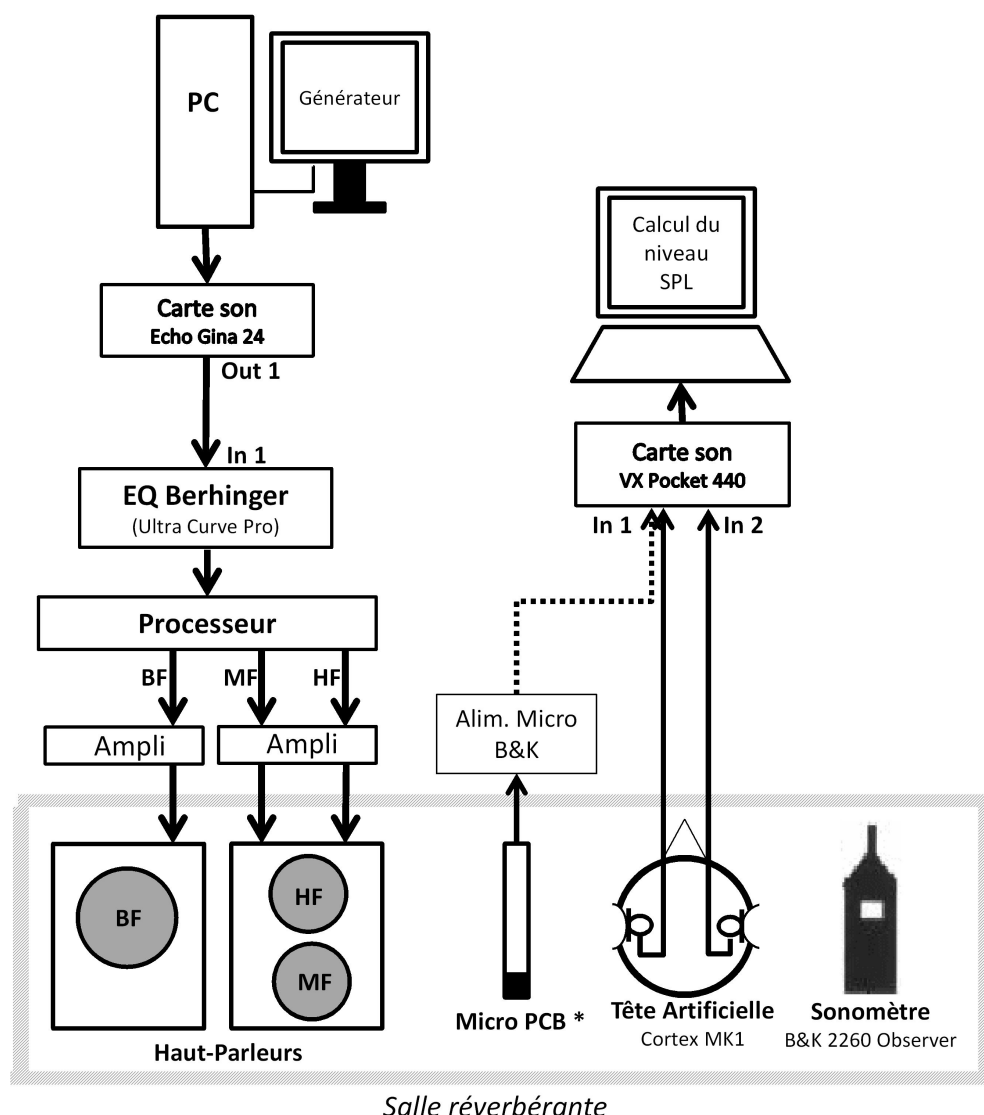


FIGURE C.1: Dispositif mis en place pour la mesure des niveaux équivalents en champ diffus. Les enceintes étaient placées en biais face à un mur (à environ 2 m) afin d'éviter au mieux le champ direct.

La première étape consistait à égaliser le système de diffusion utilisé pour la calibration. Cette égalisation a été réalisée à l'aide d'un bruit rose, par bandes de tiers d'octave, à partir de la bande centrée à 200 Hz jusqu'à la bande centrée à 8 kHz. L'erreur maximale de niveau entre deux bandes était de 1 dB (± 0.5 dB par rapport à la moyenne). Des mesures du champ acoustique au sonomètre en différents points de la salle ont permis de vérifier que l'hypothèse de champ diffus était bien respectée pour l'ensemble de la plage de fréquence.

Une deuxième étape consistait à calibrer le microphone (PCB) ainsi que la chaîne d'acquisition à l'aide d'un pistonphone (114 dB ; 1kHz), puis à ajuster le niveau de diffusion (pour le stimulus consi-

déré) jusqu'à obtenir le niveau désiré en champ diffus pour la source de référence ($L_{ref} = 65 \text{ dB SPL}$).

Le microphone était ensuite remplacé par la tête artificielle. De la même manière, chaque canal (G, D) de la tête artificielle était calibré au préalable à l'aide d'un pistonphone (114 dB ; 1kHz). Les niveaux de pression équivalents "aux tympans" (L_{eqG} et L_{eqD}) finalement mesurés sur la tête artificielle sont répertoriés dans le tableau C.1.

Fc	Micro. PCB	Tête artificielle		L_{MOY}
	L_{ref}	L_{eqG}	L_{eqD}	
1000 Hz	65.0	68.3	68.8	68.5
2236 Hz	65.0	75.5	75.8	75.7
5000 Hz	65.0	72.1	72.7	72.4

Tableau C.1: Niveaux équivalents mesurés "aux tympans" (en dB SPL) en champ diffus pour une source à 65 dB SPL. Les sources étaient des bruits à bande étroite centrés à 1 kHz, 2236 Hz et 5 kHz.

C.1.2 Calibration du casque

Une fois les niveaux acoustiques équivalents mesurés à G et à D, l'objectif était de les restituer correctement au casque.

La calibration de la tête artificielle et de la chaîne d'acquisition était réalisée à l'aide d'un pistonphone (114 dB ; 1 kHz). Le casque était préalablement placé et ajusté sur les oreilles de la tête artificielle. La source virtuelle spatialisée (par le processus de synthèse binaurale) était ensuite jouée au casque dans les mêmes conditions, strictement que celles de l'expérience. Le niveau de diffusion était enfin ajusté (indépendamment pour chaque fréquence) jusqu'à obtenir les niveaux équivalents "aux tympans" souhaités (L_{eqG} et L_{eqD}).

A titre d'exemple, les niveaux L_G et L_D , mesurés "aux tympans" à l'issue de la calibration du casque pour l'expérience du chapitre 2, sont présentés dans le tableau C.2.

Fc	Tête artificielle		L_{MOY}
	L_G	L_D	
1000 Hz	68.5	68.5	68.5
5000 Hz	72.8	71.8	72.3

Tableau C.2: Exemple de niveaux mesurés "aux tympans" (en dB SPL), lors de la restitution des sources virtuelles au casque, après calibration.

Dissymétrie G/D dans les fonctions de transfert du casque, mesure et validation de nouvelles FTCs

D.1 Contrôle de la symétrie G/D dans l'égalisation de la réponse du casque

A la suite de l'observation de la dissymétrie en sonie dans le chapitre 4, une campagne de mesures supplémentaire a été réalisée afin de contrôler la symétrie entre gauche et droite dans la correction de la réponse du casque appliquée pour la synthèse binaurale. Une évaluation de la reproductibilité des mesures a également permis de pouvoir comparer les écarts de symétrie constatés aux erreurs expérimentales pouvant être induites par le positionnement du casque sur la tête artificielle.

D.1.1 Méthode

Le son utilisé était un bruit rose, simplement convolué par les réponses impulsionnelles inverses du casque (gauche et droite) couplé avec la tête artificielle (cf. FTCIs, § B.3.1). A titre de rappel, les fonctions de transfert du casque (FTCs) correspondantes, moyennées sur trois répétitions de la mesure (avec un repositionnement du casque entre chaque mesure), sont représentées dans la figure D.1. Les courbes du bas illustrent la variabilité causée par le repositionnement du casque, à travers la différence entre les niveaux maximum et minimum obtenus pour chaque fréquence. Comme dans le processus de construction des filtres pour la synthèse, un filtrage passe-bande entre 300 Hz et 6 kHz a également été appliqué pour chaque oreille. Les signaux étaient ensuite restitués sur la tête artificielle à l'aide du casque utilisé pour les expériences de sonie. Puis les mesures "aux tympons" ont été enregistrées et analysées en fréquence par bande de tiers d'octave. Les différentes étapes de l'expérience sont représentées dans la figure D.2.

Vingt répétitions de la mesure ont été effectuées avec un repositionnement complet du casque entre chaque mesure. Le bruit était restitué à 65 dB SPL en équivalent champ diffus.

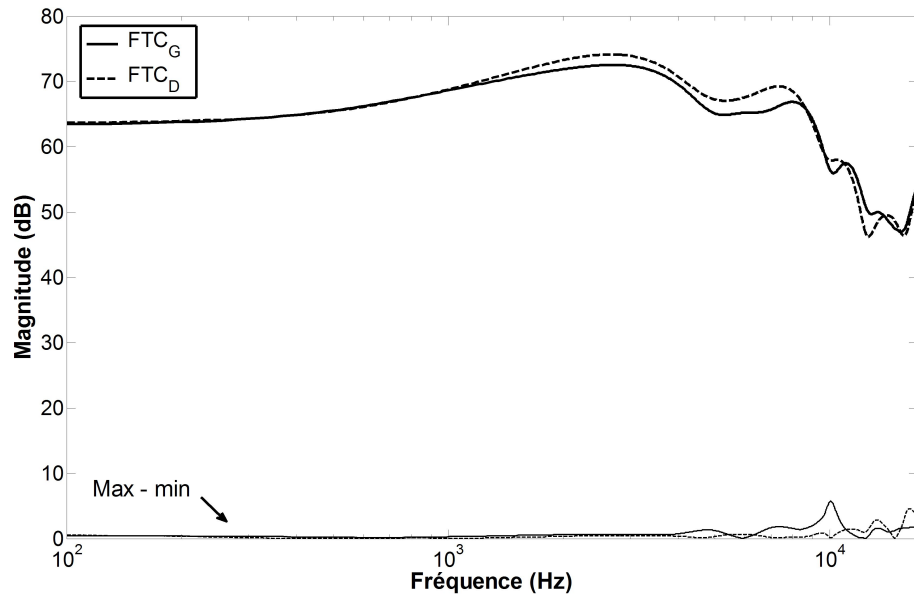


FIGURE D.1: Fonctions de transfert du casque à gauche (courbe continue) et à droite (courbe en pointillé), moyennées sur trois répétitions de la mesure avec repositionnement du casque. Les courbes du bas représentent la variabilité des mesures ($Max - Min$) pour chaque oreille.

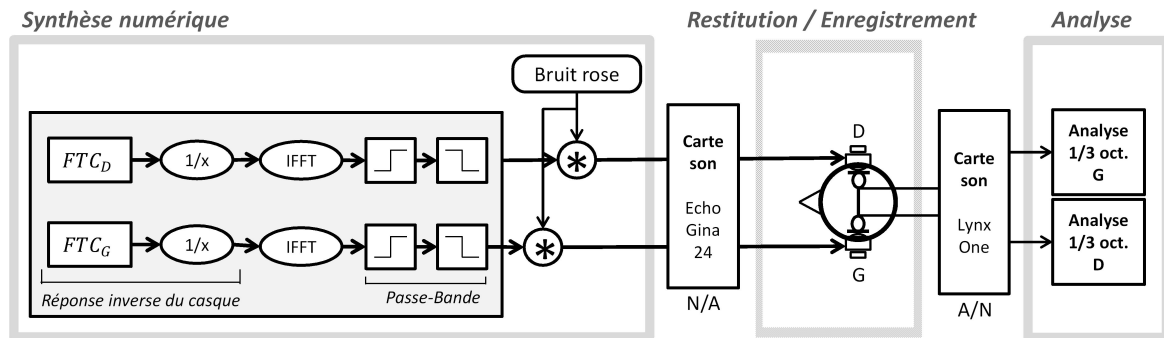


FIGURE D.2: Schéma général de l'expérience mise en place, de la synthèse des signaux à l'analyse.

D.1.2 Résultats

La figure D.3 présente les niveaux moyens (sur les vingt répétitions) par bandes de tiers d'octave, mesurés lors de la restitution du bruit rose filtré au casque pour chaque oreille. Les barres grises représentent la différence de niveau entre les deux oreilles lors de la restitution, traduisant l'ampleur de l'asymétrie.

On observe une augmentation générale de la différence de niveau entre les oreilles avec la fréquence, débutant dans le tiers d'octave centré à 1 kHz et atteignant un maximum de 1.9 dB pour $f_c = 4 \text{ kHz}$. Les écarts-types représentant la variabilité de la mesure pour chaque oreille sont faibles devant les différences interaurales de niveau à partir de 1250 Hz, ce qui montre que l'asymétrie constatée n'est pas simplement due à l'erreur expérimentale.

La figure D.1 montre qu'une différence entre les deux oreilles (avantage pour l'oreille droite)

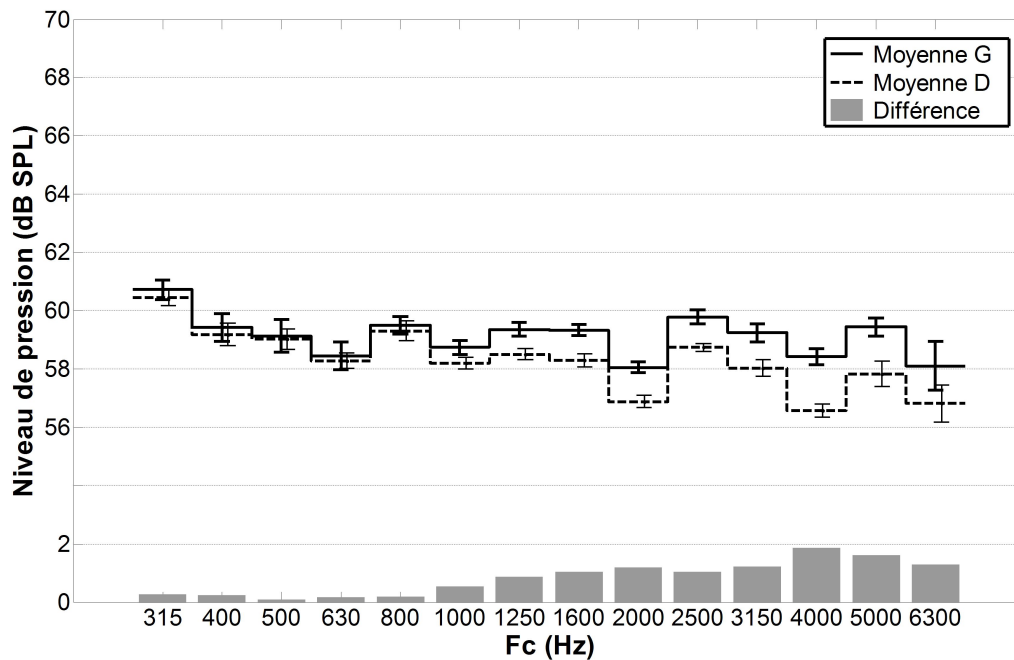


FIGURE D.3: Niveau par bande de tiers d'octave lors de la restitution du bruit rose au casque sur la tête artificielle. Les résultats sont comparés pour l'oreille G (courbe continue) et pour l'oreille D (courbe en pointillés). Les barres d'erreur représentent l'écart-type pour chaque oreille sur les 20 répétitions de la mesure, avec repositionnement du casque. Les barres grises correspondent à la différence de niveau entre les deux oreilles.

apparaissait déjà dans les fonctions de transfert, mais cette différence avait été interprétée comme une caractéristique intrinsèque du casque. D'ailleurs, on notera que l'écart maximum ne se trouvait pas exactement autour de 4 kHz (fig. D.1), mais plutôt entre 6 et 7 kHz. De plus, dans les tests mis en place pour mesurer la sonie dans le chapitre 2, les sources étaient spatialisées uniquement du côté gauche de l'auditeur. Cette dissymétrie n'avait donc toujours pas pu être détectée. En revanche, dans le protocole expérimental mis en place dans le chapitre 4, le côté de présentation de la source était un des paramètres expérimentaux. C'est ce qui a finalement permis, à travers la sous-estimation en sonie lorsque la composante à 5 kHz était spatialisée à droite, de mettre en avant la baisse de niveau du côté droit lors de la restitution. L'origine exacte de l'erreur (probablement au niveau du dispositif expérimental utilisé pour la mesure des FTCs) n'est actuellement pas connue.

Afin de préciser la dissymétrie pour les niveaux de restitutions impliqués dans les mesures de sonie (chapitres 2 et 4), la figure D.4 réunit les niveaux moyens restitués pour chaque oreille dans les bandes de tiers d'octave centrées à 1 kHz et à 5 kHz uniquement.

Cette fois-ci les barres d'erreur correspondent à l'intervalle entre les niveaux minimum et maximum mesurés sur les 20 répétitions de la mesure. A 1 kHz comme à 5 kHz, les différences obtenues entre les niveaux restitués pour l'oreille gauche et pour l'oreille droite sont significatives ($K - S; p < 0.01$). Cependant, tandis qu'à 1 kHz l'écart entre les moyennes reste faible (0.5 dB), à 5 kHz on observe un écart plus conséquent de 1.6 dB.

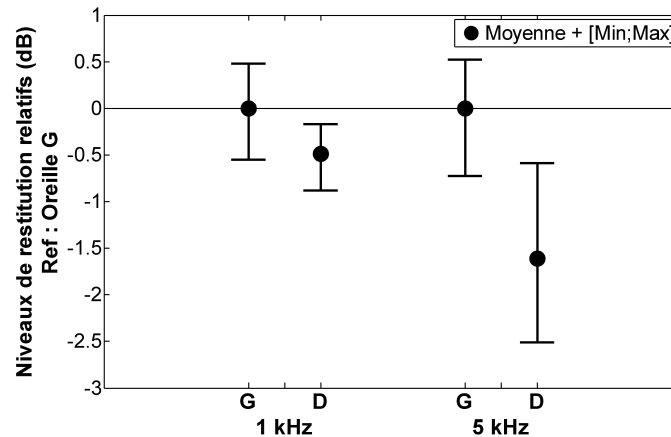


FIGURE D.4: Niveaux relatifs restitués à G et à D (par rapport aux niveaux restitués à l'oreille G) à 1 kHz et à 5 kHz. Les points et barres d'erreur représentent la moyenne et les valeurs min. et max. obtenues sur les 20 répétitions de la mesure.

D.2 Mesure et validation de nouvelles fonctions de transfert

Dans l'hypothèse où l'asymétrie constatée proviendrait des fonctions de transfert du casque initialement mesurées, de nouvelles fonctions de transfert ont été mesurées sur un plus grand nombre de répétitions (20 au lieu de 3 pour les FTCs initiales). Les FTCs moyennes sont représentées dans la figure D.5 pour chaque oreille, de manière analogue à la figure D.1.

On constate que les différences entre les deux oreilles observées dans les FTCs initiales (figure D.1) ont été fortement réduites pour les nouvelles FTCs (figure D.5). La variabilité ($Max - Min$) sur les répétitions de la mesure a légèrement augmenté, mais ceci est une conséquence principalement de l'augmentation du nombre de mesures.

Afin de contrôler l'influence de l'utilisation des nouvelles fonctions de transfert sur les niveaux restitués aux oreilles de la tête artificielle, le même protocole expérimental que celui décrit dans la partie D.1.1 a été mis en place. Les niveaux moyens obtenus par bande de tiers d'octave et pour chaque oreille sont présentés dans la figure D.6.

On constate d'une manière générale que la dissymétrie entre les deux oreilles a été fortement réduite pour l'ensemble des tiers d'octave (différence maximale d'environ 0.4 dB pour la fréquence centrale 2 kHz). L'atténuation de l'ordre de 2 dB dans la bande centrée à 6300 Hz correspond à la fin de la bande passante du bruit rose (fréquence de coupure haute du filtre à 6 kHz). Cette atténuation n'apparaissait pas clairement dans la figure D.3, ayant peut-être été compensée par une erreur dans les fonctions de transfert du casque.

Cela prouve finalement que l'asymétrie constatée dans les niveaux restitués était bien contenue dans des erreurs de mesure (ou de calcul) des fonctions de transfert du casque initiales. De nouvelles fonctions de transfert du casque ont été mesurées. L'égalisation de la réponse du casque dans le processus de restitution à partir de ces fonctions de transfert a été validée pour les bandes de tiers d'octave comprise entre $f_c = 315 \text{ Hz}$ et $f_c = 6300 \text{ Hz}$. La dernière étape consistera à vérifier si cette différence interaurale dans les niveaux restitués était bien à l'origine de l'asymétrie observée en sonie (cf. partie 4.4.3).

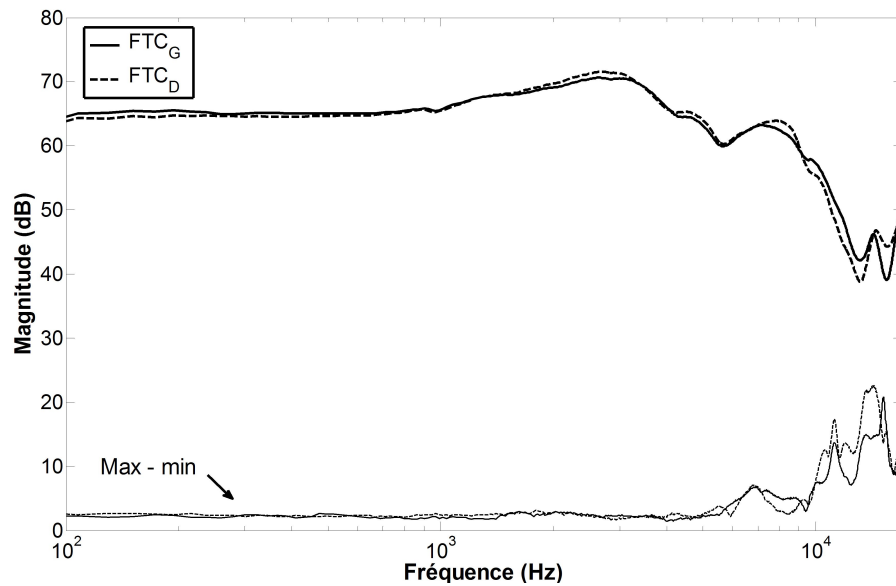


FIGURE D.5: Fonctions de transfert du casque à gauche (courbe continue) et à droite (courbe en pointillé), moyennées sur vingt répétitions de la mesure avec repositionnement du casque. Les courbes du bas représentent la variabilité des mesures (*Max – Min*) pour chaque oreille.

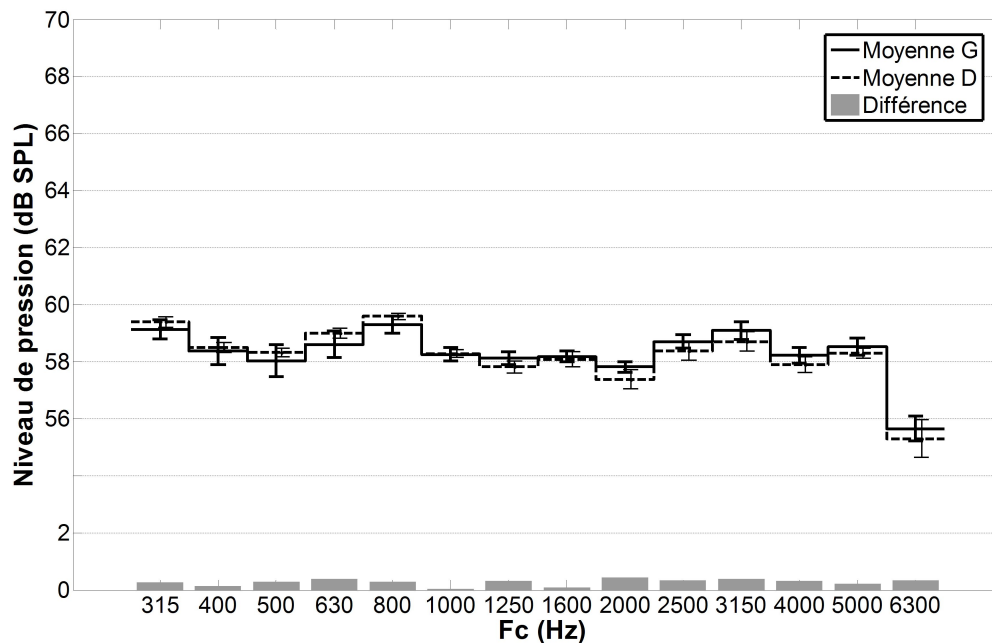


FIGURE D.6: Niveau par bande de tiers d'octave lors de la restitution . Les résultats sont comparés pour l'oreille G (courbe continue) et pour l'oreille D (courbe en pointillés). Les barres d'erreur représentent l'écart-type pour chaque oreille sur les 20 répétitions de la mesure, avec repositionnement du casque. Les barres grises correspondent à la différence de niveau entre les deux oreilles.

Analyse spectrale des sources sonores réelles

Les spectres par bandes de tiers d'octave sont présentés pour l'ensemble des sources sonores réelles utilisées dans l'expérience du chapitre 5 (A1 à A6, M1 à M6, V1, V2, S1 et S2). Chaque graphique représente le niveau de pression en fonction de la fréquence pour les canaux G (courbe bleue) et D (courbe rouge) de l'enregistrement binaural, ainsi que pour un des canaux (G ou D) du son diotique associé (courbe noire).

Le signal du son diotique pour une oreille correspondait à la moyenne des signaux temporels G et D de l'enregistrement binaural (cf. § 5.1.3). Pour les sons d'aspirateurs ou de machine à laver d'incidence non-nulle, le son diotique était calculé à partir de l'enregistrement binaural d'incidence frontale (A1 ou M1).

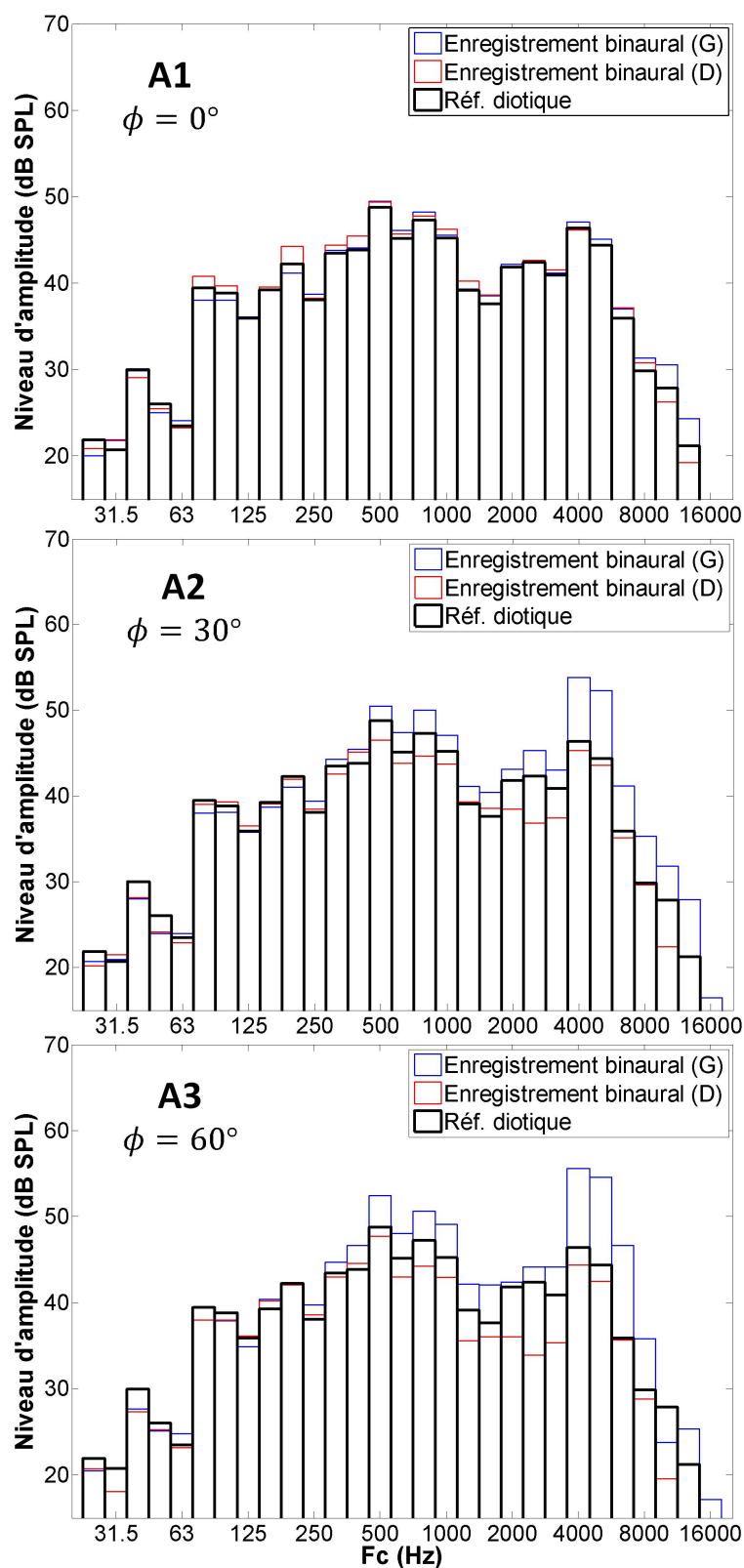


FIGURE E.1: Spectres (1/3 d'octave) pour les sons d'aspirateur à 0° , 30° et 60° : A1 (en haut), A2 (au milieu), A3 (en bas).

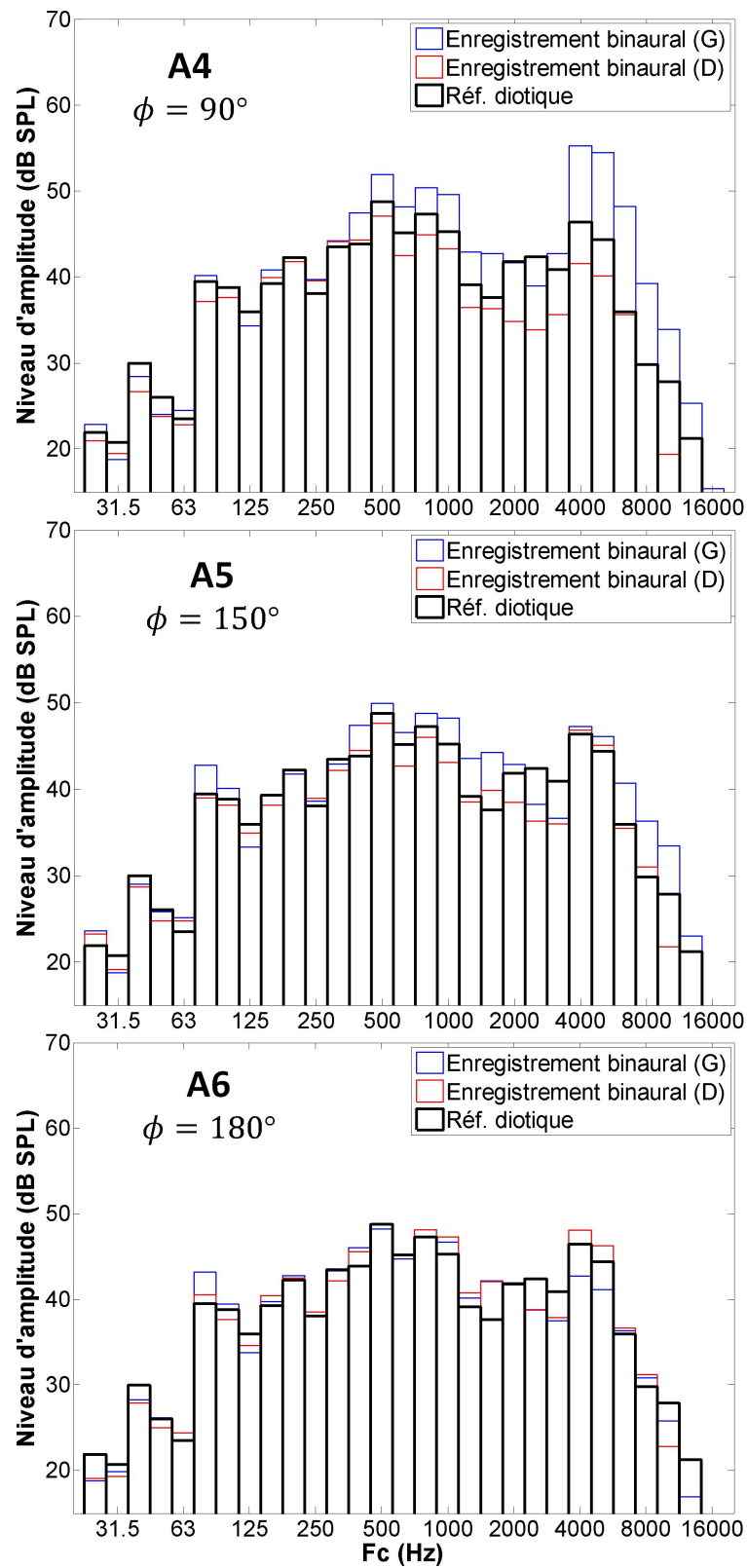


FIGURE E.2: Spectres (1/3 d'octave) pour les sons d'aspirateur à 90° , 150° et 180° : A4 (en haut), A5 (au milieu), A6 (en bas).

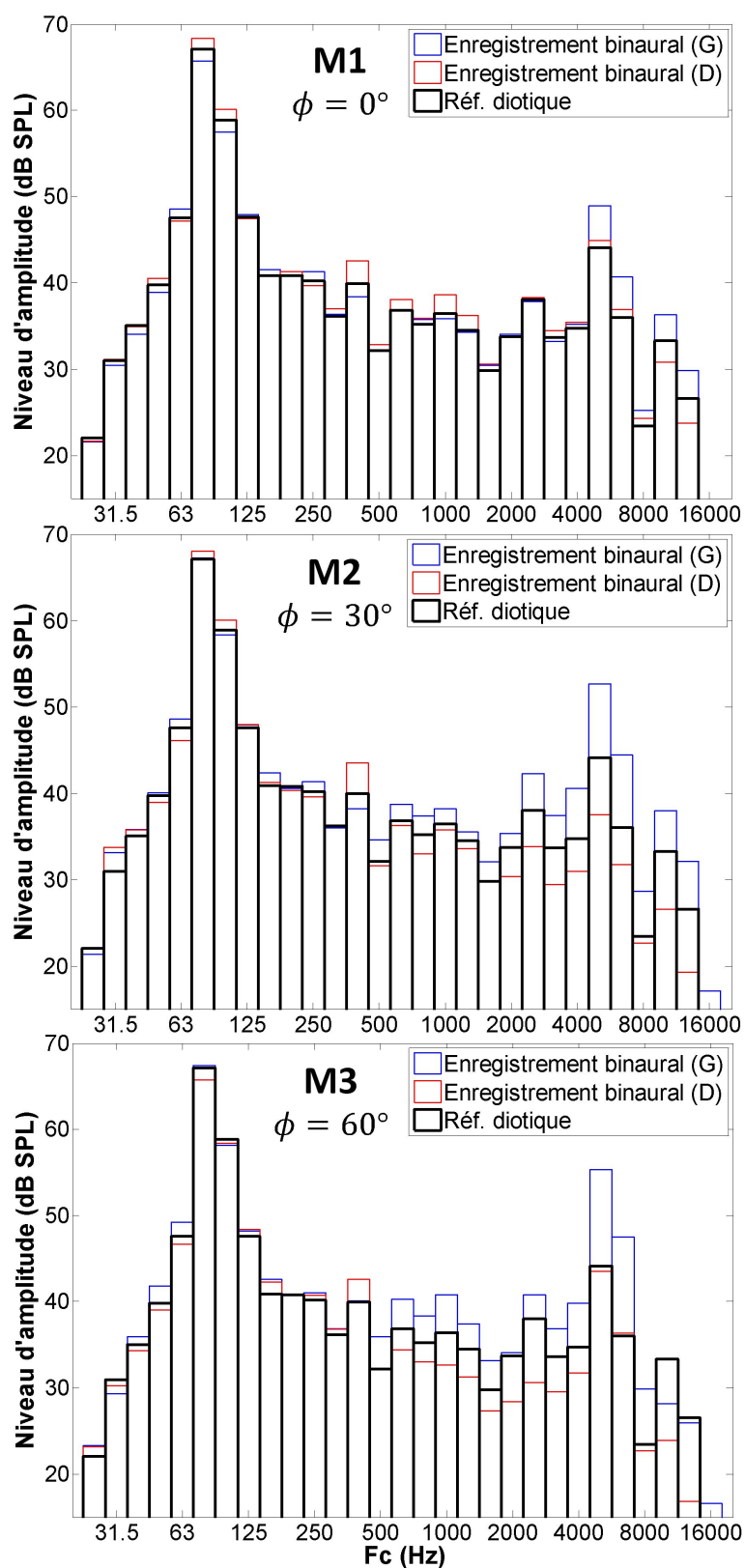


FIGURE E.3: Spectres (1/3 d'octave) pour les sons de machine à laver à 0° , 30° et 60° : M1 (en haut), M2 (au milieu), M3 (en bas).

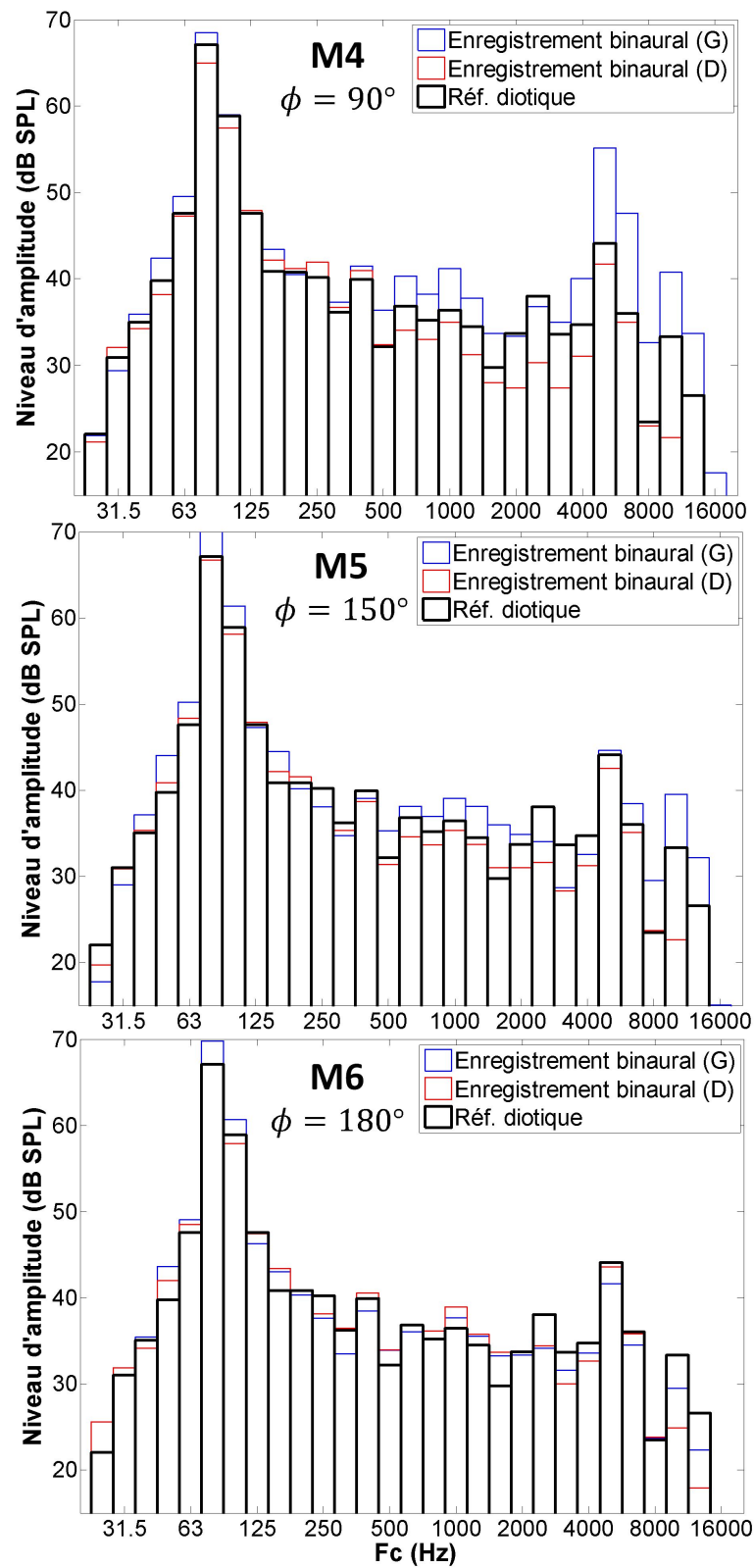


FIGURE E.4: Spectres (1/3 d'octave) pour les sons de machine à laver à 90° , 150° et 180° : M4 (en haut), M5 (au milieu), M6 (en bas).

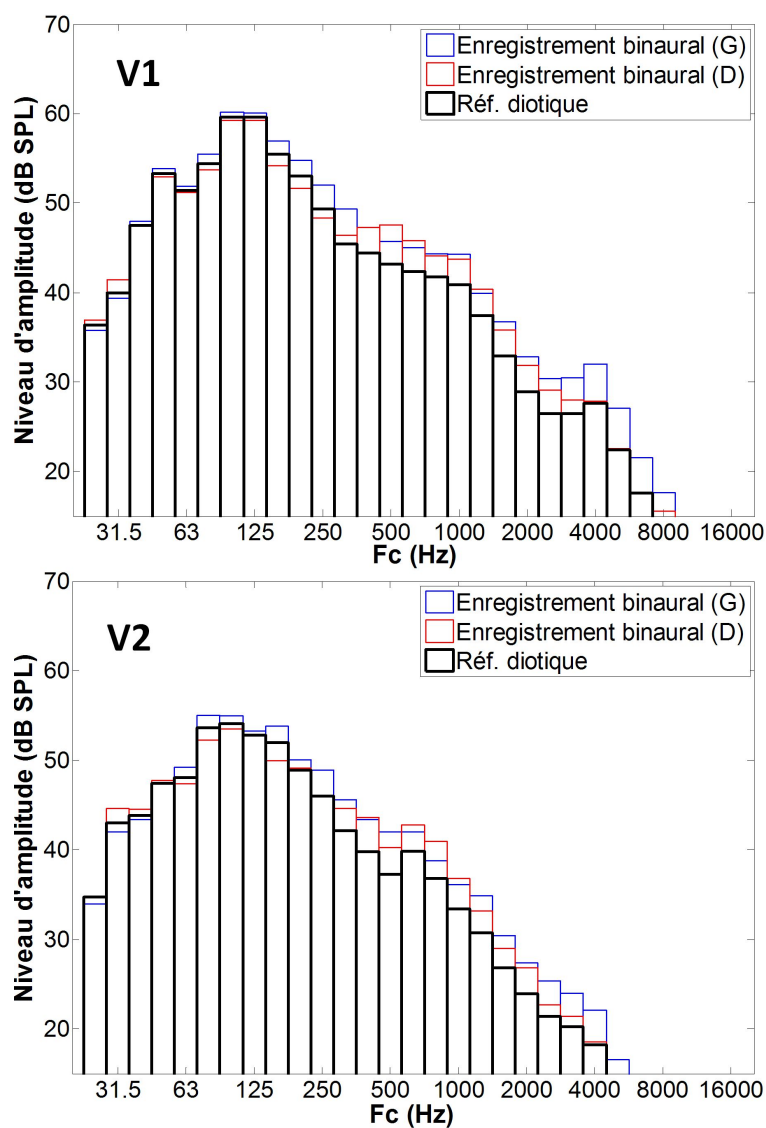


FIGURE E.5: Spectres (1/3 d'octave) pour les bruits de ventilation : V1 (en haut) et V2 (en bas).

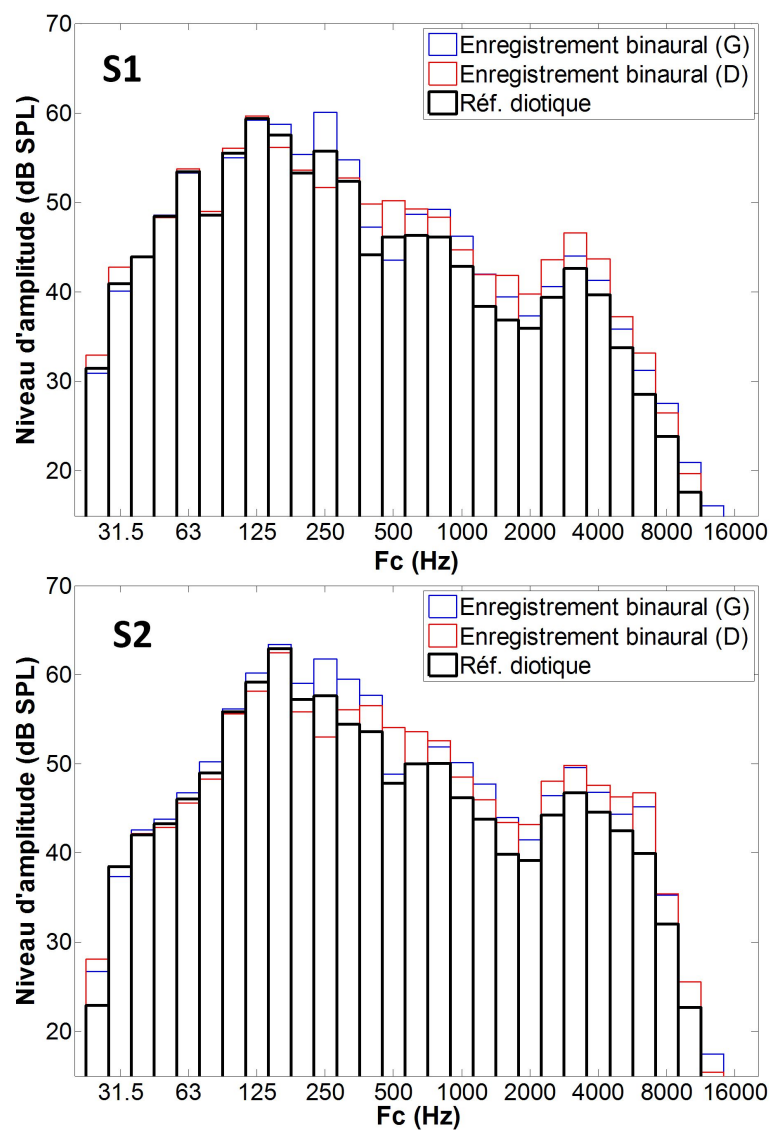


FIGURE E.6: Spectres (1/3 d'octave) pour les bruits réalisés en soufflerie : S1 (en haut) et S2 (en bas).

Références

- ALGOM, D., BEN-AHARON, B. et COHEN-RAZ, L. (1989a). Dichotic, diotic, and monaural summation of loudness : a comprehensive analysis of composition and psychophysical functions. *Perception & psychophysics*, 46(6) : 567–578. (Cité en page 34.)
- ALGOM, D., RUBIN, A. et COHEN-RAZ, L. (1989b). Binaural and temporal integration of the loudness of tones and noises. *Perception & psychophysics*, 46(2) : 155–166. (Cité en pages 13, 15 et 29.)
- ANSI-S1.11 (1986). American national standard specification for octave-band and fractional-octave-band analog and digital filters. *Acoustical Society of America*. (Cité en pages 43 et 90.)
- ANSI-S3.4 (2007). Procedure for the computation of loudness of steady sounds. *Acoustical Society of America*. (Cité en pages 1, 2, 6, 9, 23, 26, 27, 28, 29, 33, 34, 35, 36, 38, 39, 58, 60, 81, 114, 115, 116, 118, 120, 130, 131, 132, 135 et 137.)
- AYLOR, D. E. et MARKS, L. E. (1976). Perception of noise transmitted through barriers. *The Journal of the Acoustical Society of America*, 59(2) : 397–400. (Cité en page 25.)
- BEERENDS, J. G. et HOUTSMA, A. J. M. (1986). Pitch identification of simultaneous dichotic two-tone complexes. *The Journal of the Acoustical Society of America*, 80(4) : 1048–1056. (Cité en pages 102 et 103.)
- BERGLUND, B., PREIS, A. et RANKIN, K. (1990). Relationship between loudness and annoyance for ten community sounds. *Environment International*, 16(4) : 523–531. (Cité en page 1.)
- BLAUERT, J. (1997). *Spatial hearing : the psychophysics of human sound localization*. MIT press. (Cité en pages 17 et 60.)
- BOULLET, I. (2005). *La sonie des sons impulsifs : perception, mesures et modèles*. Thèse de doctorat, Université de la méditerranée Aix-Marseille II. (Cité en pages 11 et 34.)
- BOULLET, I., MAROZEAU, J. et MEUNIER, S. (2004). Temporal summation of loudness as a function of frequency and temporal pattern. In *33rd International Congress and Exposition on noise control engineering (Internoise)*, Prague, 22-25 août, Czech Republic. (Cité en page 10.)
- BREEBAART, J., VAN DE PAR, S. et KOHLRAUSCH, A. (2001). Binaural processing model based on contralateral inhibition. i. Model structure. *The Journal of the Acoustical Society of America*, 110(2) : 1074–1088. (Cité en page 30.)
- BREGMAN, A. S. (1990). *Auditory scene analysis : The perceptual organization of sound*. MIT press. (Cité en pages 25, 101 et 103.)
- BREGMAN, A. S. et TOUGAS, Y. (1989). Propagation of constraints in auditory organization. *Attention, Perception, & Psychophysics*, 46(4) : 395–396. (Cité en pages 101 et 107.)

- BUSCH-VISHNIAC, I. J., WEST, J. E., BARNHILL, C., HUNTER, T., ORELLANA, D. et CHIVUKULA, R. (2005). Noise levels in Johns Hopkins hospital. *The Journal of the Acoustical Society of America*, 118(6) : 3629–3645. (Cité en page 124.)
- BUUS, S. et FLORENTINE, M. (2002). Growth of loudness in listeners with cochlear hearing losses : recruitment reconsidered. *JARO-Journal of the Association for Research in Otolaryngology*, 3(2) : 120–139. (Cité en pages 7 et 75.)
- BUUS, S., FLORENTINE, M. et POULSEN, T. (1997). Temporal integration of loudness, loudness discrimination, and the form of the loudness function. *The Journal of the Acoustical Society of America*, 101(2) : 669–680. (Cité en pages 6, 10 et 109.)
- BUUS, S., MÜSCH, H. et FLORENTINE, M. (1998). On loudness at threshold. *The Journal of the Acoustical Society of America*, 104(1) : 399–410. (Cité en pages 6 et 109.)
- CABRERA, D., MIRANDA, L. et DASH, I. (2008). Directional loudness measurements for a multi-channel system. In *Acoustics 2008, Paris, France*. (Cité en page 18.)
- CAUSSÉ, R. et CHAVASSE, P. (1942). Difference entre l'écoute binauriculaire et monauriculaire pour la perception des intensités supraliminaire. *Comptes Rendus de la Société de Biologie*, 136 : 405–406. (Cité en page 12.)
- CHALUPPER, J. et FASTL, H. (2002). Dynamic loudness model (DLM) for normal and hearing-impaired listeners. *Acta Acustica united with Acustica*, 88(3) : 378–386. (Cité en page 1.)
- CHARBONNEAU, J., NOVAK, C. et ULE, H. (2009). Loudness prediction model comparison using the equal loudness contours. *Canadian Acoustics*, 37(3) : 64–65. (Cité en page 9.)
- CHURCHER, B., KING, A. et DAVIES, H. (1934). The measurement of noise, with special reference to engineering noise problems. *Journal of the Institution of Electrical Engineers*, 75(454) : 401–422. (Cité en page 78.)
- COLLINS, A. A. et GESCHIEDER, G. A. (1989). The measurement of loudness in individual children and adults by absolute magnitude estimation and cross-modality matching. *The Journal of the Acoustical Society of America*, 85(5) : 2012–2021. (Cité en pages 66, 72 et 73.)
- CROSS, D. V. (1973). Sequential dependencies and regression in psychophysical judgments. *Perception & Psychophysics*, 14(3) : 547–552. (Cité en pages 66, 67, 68, 69, 73 et 79.)
- CULLING, J. F. et SUMMERFIELD, Q. (1995). Perceptual separation of concurrent speech sounds : Absence of across-frequency grouping by common interaural delay. *The Journal of the Acoustical Society of America*, 98(2) : 785–797. (Cité en pages 102 et 103.)
- DARWIN, C. J. et CHRIS, J. (1997). Auditory grouping. *Trends in cognitive sciences*, 1(9) : 327–333. (Cité en page 102.)
- DARWIN, C. J. et CIOCCA, V. (1992). Grouping in pitch perception : Effects of onset asynchrony and ear of presentation of a mistuned component. *The Journal of the Acoustical Society of America*, 91(6) : 3381–3390. (Cité en pages 102 et 103.)

- DARWIN, C. J. et HUKIN, R. (1999). Auditory objects of attention : the role of interaural time differences. *Journal of Experimental Psychology : Human perception and performance*, 25(3) : 617. (Cité en page 101.)
- DARWIN, C. J. et HUKIN, R. W. (1997). Perceptual segregation of a harmonic from a vowel by interaural time difference and frequency proximity. *The Journal of the Acoustical Society of America*, 102(4) : 2316–2324. (Cité en page 102.)
- DAVID, M., LAVANDIER, M. et GRIMAULT, N. (2014). Room and head coloration can induce obligatory stream segregation. *The Journal of the Acoustical Society of America*, 136(1) : 5–8. (Cité en page 101.)
- DECARLO, L. T. (2003). An application of a dynamic model of judgment to magnitude production. *Perception & Psychophysics*, 65(1) : 152–162. (Cité en pages 73, 79 et 80.)
- DECARLO, L. T. (2005). On bias in magnitude scaling and some conjectures of Stevens. *Perception & psychophysics*, 67(5) : 886–896. (Cité en pages 13, 73, 79 et 80.)
- DIN-45631 (1990). Berechnung des laustärkepegels und der lautheit aus dem gerauschkpektrum. *Verfahren nach E. Zwicker*. (Cité en pages 1, 9, 26, 33, 34, 118 et 135.)
- DUBROVSKII, N. A., CHERNYAK, R. I. et SHAPIRO, V. M. (1972). Binaural summation of differently correlated noises. *Soviet Physics Acoustics*, 17 : 468–473. (Cité en pages 16 et 132.)
- EDMONDS, B. A. et CULLING, J. F. (2009). Interaural correlation and the binaural summation of loudness. *The Journal of the Acoustical Society of America*, 125(6) : 3865–3870. (Cité en pages 16, 19, 23, 35, 132, 133, 134, 137 et 138.)
- EICHENLAUB, C., CHOUARD, N. et WEBER, R. (1996). On the influence of interaural correlation on binaural loudness for broadband noise. In *Proceedings of Internoise '96*, pages 2227–2230. (Cité en pages 16 et 132.)
- EKMAN, G., BERGLUND, B. et BERGLUND, U. (1966). Loudness as a function of the duration of auditory stimulation. *Scandinavian journal of psychology*, 7(1) : 201–208. (Cité en page 10.)
- ELLERMEIER, W. et FAULHAMMER, G. (2000). Empirical evaluation of axioms fundamental to Stevens's ratio-scaling approach : I. Loudness production. *Perception & Psychophysics*, 62(8) : 1505–1511. (Cité en pages 13 et 73.)
- EMMERICH, D. S., HARRIS, J., BROWN, W. S. et SPRINGER, S. P. (1988). The relationship between auditory sensitivity and ear asymmetry on a dichotic listening task. *Neuropsychologia*, 26(1) : 133–143. (Cité en page 70.)
- EPSTEIN, M. et FLORENTINE, M. (2005). A test of the equal-loudness-ratio hypothesis using cross-modality matching functions. *The Journal of the Acoustical Society of America*, 118(2) : 907–913. (Cité en pages 6, 10 et 14.)
- EPSTEIN, M. et FLORENTINE, M. (2009). Binaural loudness summation for speech and tones presented via earphones and loudspeakers. *Ear and hearing*, 30(2) : 234–237. (Cité en page 25.)

- EPSTEIN, M. et FLORENTINE, M. (2012). Binaural loudness summation for speech presented via earphones and loudspeaker with and without visual cues. *The Journal of the Acoustical Society of America*, 131(5) : 3981–3988. (Cité en page 25.)
- FARINA, A. (2000). Simultaneous measurement of impulse response and distortion with a swept-sine technique. In *Audio Engineering Society Convention 108*. Audio Engineering Society. (Cité en page 147.)
- FASTL, H. (2004). Audio-visual interactions in loudness evaluation. In *Proceedings of the 18th International Congress on Acoustics*, pages 1161–1166. (Cité en pages 25 et 35.)
- FASTL, H. et FLORENTINE, M. (2011). Loudness in daily environments. In *Loudness*, pages 199–221. Springer. (Cité en page 25.)
- FASTL, H., VÖLK, F. et STRAUBINGER, M. (2009). Standards for calculating loudness of stationary or time-varying sounds. In *Proceedings of Internoise, Ottawa*. (Cité en page 33.)
- FLETCHER, H. et MUNSON, W. A. (1933). Loudness, its definition, measurement and calculation. *The journal of the acoustical society of America*, 5(2) : 82–108. (Cité en pages 8, 11, 12, 22, 26 et 29.)
- FLORENTINE, M., BONDING, P. *et al.* (1978). Loudness of complex sounds as a function of the standard stimulus and the number of components. *The Journal of the Acoustical Society of America*, 64(4) : 1036–1040. (Cité en pages 10 et 106.)
- FLORENTINE, M., BUUS, S. et POULSEN, T. (1996). Temporal integration of loudness as a function of level. *The Journal of the Acoustical Society of America*, 99(3) : 1633–1644. (Cité en pages 6, 7 et 10.)
- FLORENTINE, M. et EPSTEIN, M. (2006). To honor Stevens and repeal his law (for the auditory system). *Fechner Day*, pages 37–42. (Cité en pages 7 et 75.)
- GARDNER, B., MARTIN, K. *et al.* (1994). HRTF measurements of a KEMAR dummy-head microphone. *Vision and Modeling Group, Media Laboratory, Massachusetts Institute of Technology*. (Cité en pages 51 et 149.)
- GEIGER, P. et FIRESTONE, F. (1933). The estimation of fractional loudness. *The Journal of the Acoustical Society of America*, 5(1) : 25–30. (Cité en pages 73 et 77.)
- GESCHIEDER, G. A. et NIBLETTE, R. K. (1967). Cross-modality masking for touch and hearing. *Journal of experimental psychology*, 74(3) : 313. (Cité en page 25.)
- GIGERENZER, G. et STRUBE, G. (1983). Are there limits to binaural additivity of loudness ? *J. Exp. Psychol. Hum. Percept. and Perform.*, 9(1) : 126–136. (Cité en pages 2 et 30.)
- GILLMEISTER, H. et EIMER, M. (2007). Tactile enhancement of auditory detection and perceived loudness. *Brain Research*, 1160 : 58 – 68. (Cité en page 25.)
- GLASBERG, B. R. et MOORE, B. C. (1990). Derivation of auditory filter shapes from notched-noise data. *Hearing research*, 47(1) : 103–138. (Cité en pages 10 et 28.)

- GLASBERG, B. R. et MOORE, B. C. (2002). A model of loudness applicable to time-varying sounds. *Journal of the Audio Engineering Society*, 50(5) : 331–342. (Cité en page 1.)
- GREEN, D. M., LUCE, R. D. et DUNCAN, J. E. (1977). Variability and sequential effects in magnitude production and estimation of auditory intensity. *Perception & Psychophysics*, 22(5) : 450–456. (Cité en page 79.)
- GREENWOOD, D. D. (1961). Critical bandwidth and the frequency coordinates of the basilar membrane. *The Journal of the Acoustical Society of America*, 33(10) : 1344–1356. (Cité en page 10.)
- GRIMAULT, N., MCADAMS, S. et ALLEN, J. (2007). Auditory scene analysis : A prerequisite for loudness perception. In *Hearing - From Sensory Processing to Perception*, pages 295–302. Springer Berlin Heidelberg. (Cité en pages 25, 26, 107 et 138.)
- HAWKSFORD, M. J. (1997). Digital signal processing tools for loudspeaker evaluation and discrete-time crossover design. *Journal of the Audio Engineering Society*, 45(1/2) : 37–62. (Cité en page 150.)
- HELLMAN, R. P. (1981). Stability of individual loudness functions obtained by magnitude estimation and production. *Perception & Psychophysics*, 29(1) : 63–70. (Cité en pages 73, 76 et 79.)
- HELLMAN, R. P. et MEISELMAN, C. H. (1988). Prediction of individual loudness exponents from cross-modality matching. *Journal of Speech, Language and Hearing Research*, 31(4) : 605–615. (Cité en pages 7, 62, 73, 76, 80 et 85.)
- HELLMAN, R. P. et MEISELMAN, C. H. (1993). Rate of loudness growth for pure tones in normal and impaired hearing. *The Journal of the Acoustical Society of America*, 93(2) : 966–975. (Cité en pages 7, 66 et 72.)
- HELLMAN, R. P., MIŚKIEWICZ, A. et SCHARF, B. (1997). Loudness adaptation and excitation patterns : Effects of frequency and level. *The Journal of the Acoustical Society of America*, 101(4) : 2176–2185. (Cité en page 11.)
- HELLMAN, R. P. et ZWISLOCKI, J. (1963). Monaural loudness function at 1000 cps and interaural summation. *The Journal of the Acoustical Society of America*, 35(6) : 856–865. (Cité en pages 2, 13 et 29.)
- HELLMAN, R. P. et ZWISLOCKI, J. (1968). Loudness determination at low sound frequencies. *The Journal of the Acoustical Society of America*, 43(1) : 60–64. (Cité en page 72.)
- HELLSTRÖM, Å. (1979). Time errors and differential sensation weighting. *Journal of Experimental Psychology : Human Perception and Performance*, 5(3) : 460. (Cité en pages 44, 92 et 125.)
- HIRSH, I. J. (1948). The influence of interaural phase on interaural summation and inhibition. *The Journal of the Acoustical Society of America*, 20(4) : 536–544. (Cité en page 15.)
- HIRVONEN, T. (2005). Segregation of two simultaneously arriving narrowband noise signals as a function of spatial and frequency separation. In *8th International Conference on Digital Audio Effects, (DAFx'05), Madrid, Spain*. (Cité en pages 102, 103 et 107.)

- HUKIN, R. et DARWIN, C. (1995). Effects of contralateral presentation and of interaural time differences in segregating a harmonic from a vowel. *The Journal of the Acoustical Society of America*, 98(3) : 1380–1387. (Cité en page 102.)
- IH, J.-G. et CHA, J.-H. (2003). A loudness model in dichotic conditions. *The Journal of the Acoustical Society of America*, 114(4) : 2352–2352. (Cité en page 15.)
- INDOW, T. et STEVENS, S. (1966). Scaling of saturation and hue. *Perception & Psychophysics*, 1(2) : 253–271. (Cité en page 80.)
- IRWIN, R. (1965). Binaural summation of thermal noises of equal and unequal power in each ear. *The American journal of psychology*, 78(1) : 57–65. (Cité en page 12.)
- ISO-226 (1987). Normal equal-loudness level contours. *International Organization for Standardization*. (Cité en pages 8 et 9.)
- ISO-226 (2003). Normal equal-loudness level contours. *International Organization for Standardization*. (Cité en pages 6, 8, 9, 91, 105, 108 et 111.)
- ISO-532B (1975). Méthode de calcul du niveau d'isophonie. *International Organization for Standardization*. (Cité en pages 1, 6, 9, 20, 21, 23, 26, 27, 28, 33, 35, 39, 112, 118 et 135.)
- ISO11904-2 (2004). Acoustics. Determination of sound immission from sound sources placed close to the ear - part 2 : Technique using a manikin. *International Organization for Standardization, Geneva, Switzerland*. (Cité en page 22.)
- ISO389-7 (2005). Acoustics. Reference zero for the calibration of audiometric equipment. Reference threshold of hearing under free-field and diffuse-field listening conditions. *International Organization for Standardization, Geneva, Switzerland*. (Cité en page 23.)
- ISO/R-131 (1959). Expression of physical and subjective magnitudes of sound. *International Organization for Standardization, Geneva*. (Cité en page 80.)
- ISO/R-226 (1961). Normal equal-loudness contours for pure tones and normal threshold of hearing under free field listening conditions. *International Organization for Standardization, Switzerland*. (Cité en pages 8 et 9.)
- KEEN, K. (1972). Preservation of constant loudness with interaural amplitude asymmetry. *The Journal of the Acoustical Society of America*, 52(4) : 1193–1196. (Cité en pages 15 et 34.)
- KIDD, G., MASON, C. R., BRUGHERA, A. et HARTMANN, W. M. (2005). The role of reverberation in release from masking due to spatial separation of sources for speech identification. *Acta acustica united with acustica*, 91(3) : 526–536. (Cité en page 101.)
- KOEHL, V. et PAQUIER, M. (2014). Influence de différences interaurales de temps sur la sonie de sons purs en basse fréquence. *Actes du 12e Congrès Français d'Acoustique*, pages 1947–1953. (Cité en page 16.)

- KOEHL, V. et PAQUIER, M. (2015). Loudness of low-frequency pure tones lateralized by interaural time differences. *The Journal of the Acoustical Society of America*, 137(2) : 1040–1043. (Cité en page 16.)
- KUHL, W. et WESTPHAL, W. (1959). Differences in loudness levels between a plane wave and a diffuse sound field. *Acustica*, 9 : 407–408. (Cité en page 22.)
- LEIBOLD, L. J., TAN, H., KHADDAM, S. et JESTEADT, W. (2007). Contributions of individual components to the overall loudness of a multitone complex. *The Journal of the Acoustical Society of America*, 121(5) : 2822–2831. (Cité en page 10.)
- LEVITT, H. (1971). Transformed up-down methods in psychoacoustics. *The Journal of the Acoustical Society of America*, 49(2) : 467–477. (Cité en page 44.)
- LOCHNER, J. et BURGER, J. (1961). Form of the loudness function in the presence of a masking noise. *Journal of the Acoustical Society of America*, 33(12) : 1705–1707. (Cité en page 6.)
- LOGUE, A. (1976). Individual differences in magnitude estimation of loudness. *Perception & Psychophysics*, 19(3) : 279–280. (Cité en page 66.)
- MACPHERSON, E. A. (1995). A review of auditory perceptual theories and the prospects for an ecological account. Dissertation. (Cité en page 24.)
- MACPHERSON, E. A. et MIDDLEBROOKS, J. C. (2002). Listener weighting of cues for lateral angle : The duplex theory of sound localization revisited. *The Journal of the Acoustical Society of America*, 111(5) : 2219–2236. (Cité en page 102.)
- MARKS, L. E. (1974). On scales of sensation : Prolegomena to any future psychophysics that will be able to come forth as science. *Perception & Psychophysics*, 16(2) : 358–376. (Cité en pages 76, 79 et 80.)
- MARKS, L. E. (1978). Binaural summation of the loudness of pure tones. *The Journal of the Acoustical Society of America*, 64(1) : 107–113. (Cité en pages 2, 11, 12, 13, 15, 29, 34, 61 et 69.)
- MARKS, L. E. (1980). Binaural summation of loudness : Noise and two-tone complexes. *Perception & Psychophysics*, 27(6) : 489–498. (Cité en pages 13, 15 et 69.)
- MAROZEAU, J., EPSTEIN, M., FLORENTINE, M. et DALEY, B. (2006). A test of the binaural equal-loudness-ratio hypothesis for tones. *The Journal of the Acoustical Society of America*, 120(6) : 3870–3877. (Cité en pages 2, 7, 12, 13, 14, 15 et 61.)
- MAROZEAU, J. et FLORENTINE, M. (2009). Testing the binaural equal-loudness-ratio hypothesis with hearing-impaired listeners. *The Journal of the Acoustical Society of America*, 126(1) : 310–317. (Cité en pages 69 et 72.)
- MATELJAN, I. et UGRINOVIĆ, K. (2003). The comparison of room impulse response measuring systems. In *Proceedings of the First Congress of Alps Adria Acoustics Association*. (Cité en page 147.)

- MCADAMS, S., BOTTE, M.-C. et DRAKE, C. (1998). Auditory continuity and loudness computation. *The Journal of the Acoustical Society of America*, 103(3) : 1580–1591. (Cité en pages 25, 26, 35 et 138.)
- MEUNIER, S. et MARCHIONI, A. (2002). Loudness of sounds with temporal variable intensity. *In Forum Acusticum 2002 -European and Japanese Symposium, Sevilla, Spain*. (Cité en page 1.)
- MEUNIER, S., MARCHIONI, A. et RABAU, G. (2000). Subjective evaluation of loudness models using synthesized and environmental sounds. *Proceeding of 29th Internoise*. (Cité en pages 34 et 118.)
- MIDDLEBROOKS, J. C. et ONSAN, Z. A. (2012). Stream segregation with high spatial acuity. *The Journal of the Acoustical Society of America*, 132(6) : 3896–3911. (Cité en page 101.)
- MIRANDA, L. et CABRERA, D. (2008). Evaluation of binaural loudness models with signals of different diffusivity. *In Proceedings of the Australian Acoustical Society Conference (Acoustics 2008), Geelong, Australia*, pages 24–26. (Cité en pages 23, 35, 132, 133, 134 et 138.)
- MISKIEWICZ, A., SCHARF, B., HELLMAN, R. et MEISELMAN, C. (1993). Loudness adaptation at high frequencies. *The Journal of the Acoustical Society of America*, 94(3) : 1281–1286. (Cité en page 11.)
- MOLLA, S., BOULLET, I., MEUNIER, S., RABAU, G., GAUDUIN, B., BOUSSARD, P. *et al.* (2010). Calcul des indicateurs de sonie : revue des algorithmes et implémentation. *10ème Congrès Français d'Acoustique*. (Cité en pages 26 et 27.)
- MØLLER, H. (1992). Fundamentals of binaural technology. *Applied acoustics*, 36(3) : 171–218. (Cité en pages 17, 146 et 149.)
- MOORE, B. C. et GLASBERG, B. R. (1983). Suggested formulae for calculating auditory-filter bandwidths and excitation patterns. *The Journal of the Acoustical Society of America*, 74(3) : 750–753. (Cité en page 10.)
- MOORE, B. C. et GLASBERG, B. R. (1996). A revision of zwicker's loudness model. *Acta Acustica united with Acustica*, 82(2) : 335–345. (Cité en page 118.)
- MOORE, B. C. et GLASBERG, B. R. (2007). Modeling binaural loudness. *The Journal of the Acoustical Society of America*, 121(3) : 1604–1612. (Cité en pages ix, 2, 29, 30, 32, 33, 34, 35, 58, 61, 81, 114, 115, 116, 120, 130, 131, 134, 135 et 137.)
- MOORE, B. C., GLASBERG, B. R. et BAER, T. (1997). A model for the prediction of thresholds, loudness, and partial loudness. *Journal of the Audio Engineering Society*, 45(4) : 224–240. (Cité en pages 10, 26, 29, 30, 34, 35 et 118.)
- MULLIGAN, B., GOODMAN, L., GLEISNER, D. et FAUPEL, M. (1985). Steps in loudness summation. *The Journal of the Acoustical Society of America*, 77(3) : 1141–1154. (Cité en pages 15, 16 et 35.)
- ODGAARD, E. C., ARIEH, Y. et MARKS, L. E. (2004). Brighter noise : sensory enhancement of perceived loudness by concurrent visual stimulation. *Cognitive, Affective, & Behavioral Neuroscience*, 4(2) : 127–132. (Cité en page 24.)

- PARIZET, E. et KOEHL, V. (2011). Influence of train colour on loudness judgments. *Acta Acustica united with Acustica*, 97(2) : 347–349. (Cité en page 25.)
- PATSOURAS, C., FILIPPOU, T. G. et FASTL, H. (2002). Influences of color on the loudness judgement. In *Proc. Forum Acusticum, Sevilla*, volume 2. (Cité en page 25.)
- PETERSEN, J. (1990). Estimation of loudness and apparent distance of pure tones in a free field. *Acta Acustica united with Acustica*, 70(1) : 61–65. (Cité en page 24.)
- PIN, M. C. et CHOCHOLLE, R. (1970). L'effet du déphasage interaural sur la sonie binaurale. *Soc. Biol.*, 164 : 76–81. (Cité en page 15.)
- PIN, M. C. et CHOCHOLLE, R. (1971). L'effet du déphasage interaural sur la sonie binaurale d'un son de 400 Hz ; ses rapports avec les battements binauraux. *Soc. Biol.*, 165 : 41–47. (Cité en page 15.)
- PONSOT, E., SUSINI, P. et MEUNIER, S. (2015). A robust asymmetry in loudness between rising- and falling-intensity tones. *Attention, Perception, & Psychophysics*, pages 1–14. (Cité en page 1.)
- POULSEN, T. et HAN, L. A. (2000). The binaural free field hearing threshold for pure tones from 125 Hz to 16 kHz. *Acta Acustica united with Acustica*, 86(2) : 333–337. (Cité en page 106.)
- RADER, T., MORINAGA, M., MATSUI, T., FASTL, H., KUWANO, S. et NAMBA, S. (2004). Cross-cultural effects in audio-visual interactions. *Transactions of the TC Noise and Vibration of Acoust Soc Jpn*, N-2004-31. (Cité en page 25.)
- REYNOLDS, G. S. et STEVENS, S. (1960). Binaural summation of loudness. *The Journal of the Acoustical Society of America*, 32(10) : 1337–1344. (Cité en page 12.)
- ROBINSON, D. W. (1957). The subjective loudness scale. *Acustica*, 7 : 217–233. (Cité en page 78.)
- ROBINSON, D. W. et DADSON, R. S. (1956). A re-determination of the equal-loudness relations for pure tones. *British Journal of Applied Physics*, 7(5) : 166–181. (Cité en pages 8 et 9.)
- ROBINSON, D. W. et DADSON, R. S. (1957). Threshold of hearing and equal-loudness relations for pure tones, and the loudness function. *The Journal of the Acoustical Society of America*, 29(12) : 1284–1288. (Cité en page 106.)
- ROBINSON, D. W. et WHITTLE, L. S. (1960). The loudness of directional sound fields. *Acustica*, 10 : 74–80. (Cité en pages 2, 17, 18, 54, 55, 58, 60 et 61.)
- ROBINSON, D. W., WHITTLE, L. S. et BOWSER, J. M. (1961). The loudness of diffuse sound fields. *Acustica*, 11 : 397–404. (Cité en page 22.)
- SCHARF, B. (1969). Dichotic summation of loudness. *The Journal of the Acoustical Society of America*, 45(5) : 1193–1205. (Cité en pages 16, 29, 30, 34, 106, 107 et 111.)
- SCHARF, B. (1974). Loudness summation between tones from two loudspeakers. *The Journal of the Acoustical Society of America*, 56(2) : 589–593. (Cité en page 20.)

- SCHARF, B. (1978). Loudness. *Handbook of perception*, 4 : 187–242. (Cité en page 6.)
- SCHARF, B. et FISHKEN, D. (1970). Binaural summation of loudness : Reconsidered. *Journal of Experimental Psychology*, 86(3) : 374. (Cité en pages 2, 12 et 13.)
- SCHLITTENLACHER, J., ELLERMEIER, W. et ARSENEAU, J. (2014). Binaural loudness gain measured by simple reaction time. *Attention, Perception, & Psychophysics*, pages 1–8. (Cité en page 15.)
- SCHLITTENLACHER, J., ELLERMEIER, W. et HASHIMOTO, T. (2015). Spectral loudness summation : Shortcomings of current standards. *The Journal of the Acoustical Society of America*, 137(1) : EL26–EL31. (Cité en pages 34, 118, 120 et 137.)
- SCHLITTENLACHER, J., HASHIMOTO, T., FASTL, H., NAMBA, S., KUWANO, S. et HATANO, S. (2011). Loudness of pink noise and stationary technical sounds. *In Proceedings of InterNoise*. (Cité en page 33.)
- SCHNEIDER, B. A. et COHEN, A. J. (1997). Binaural additivity of loudness in children and adults. *Perception & psychophysics*, 59(5) : 655–664. (Cité en page 70.)
- SCHNEIDER, B. A., WRIGHT, A. A., EDELHEIT, W., HOCK, P. et HUMPHREY, C. (1972). Equal loudness contours derived from sensory magnitude judgments. *The Journal of the Acoustical Society of America*, 51(6B) : 1951–1959. (Cité en page 72.)
- SCHWARTZ, A., MCDERMOTT, J. H. et SHINN-CUNNINGHAM, B. (2012). Spatial cues alone produce inaccurate sound segregation : The effect of interaural time differences. *The Journal of the Acoustical Society of America*, 132(1) : 357–368. (Cité en page 102.)
- SIVONEN, V. P. (2007). Directional loudness and binaural summation for wideband and reverberant sounds. *The Journal of the Acoustical Society of America*, 121(5) : 2852–2861. (Cité en pages 2, 18, 22, 24, 32, 34, 43, 44 et 54.)
- SIVONEN, V. P. et ELLERMEIER, W. (2006). Directional loudness in an anechoic sound field, head-related transfer functions, and binaural summation. *The Journal of the Acoustical Society of America*, 119(5) : 2965–2980. (Cité en pages 1, 2, 18, 19, 20, 32, 34, 43, 44, 46, 52, 54, 60, 61 et 126.)
- SIVONEN, V. P. et ELLERMEIER, W. (2008). Binaural loudness for artificial-head measurements in directional sound fields. *Journal of the Audio Engineering Society*, 56(6) : 452–461. (Cité en pages ix, xi, 2, 3, 18, 19, 20, 21, 22, 29, 32, 33, 34, 35, 39, 43, 44, 48, 51, 54, 58, 60, 61, 62, 63, 81, 91, 94, 98, 112, 113, 114, 115, 116, 118, 120, 130, 131, 132, 134, 135, 136, 137, 145 et 149.)
- SIVONEN, V. P. et ELLERMEIER, W. (2011). Binaural loudness. *In Loudness*, pages 169–197. Springer. (Cité en pages 11 et 22.)
- SIVONEN, V. P., MINNAAR, P. et ELLERMEIER, W. (2005). Effect of direction on loudness in individual binaural synthesis. *In Audio Engineering Society Convention 118*. (Cité en pages 2, 18, 32, 34, 145 et 154.)
- SLABBEKOORN, H. et PEET, M. (2003). Ecology : Birds sing at a higher pitch in urban noise. *Nature*, 424(6946) : 267–267. (Cité en page 124.)

- SMALL JR, A. M., BRANDT, J. F. et COX, P. G. (1962). Loudness as a function of signal duration. *The Journal of the Acoustical Society of America*, 34(4) : 513–514. (Cité en page 10.)
- SONG, W. (2007). *Beamforming applied to psychoacoustics*. Thèse de doctorat, Sound Quality Research Unit (SQRU) Section of Acoustics, Department of Electronic Systems Aalborg University, Denmark. (Cité en pages 3, 20, 21, 89, 97, 104, 118, 119, 120 et 137.)
- SONG, W., ELLERMEIER, W. et MINNAAR, P. (2005). Perceived loudness of spatially distributed sound sources. In *Proceedings of Forum Acusticum, Budapest, Hungary*, pages 1665–1670. (Cité en page 34.)
- STAN, G.-B., EMBRECHTS, J.-J. et ARCHAMBEAU, D. (2002). Comparison of different impulse response measurement techniques. *Journal of the Audio Engineering Society*, 50(4) : 249–262. (Cité en page 147.)
- STEPHENS, S. (1974). Methodological factors influencing loudness of short duration sounds. *Journal of Sound and Vibration*, 37(2) : 235–246. (Cité en page 10.)
- STEVENS, S. (1955). The measurement of loudness. *The Journal of the Acoustical Society of America*, 27(5) : 815–829. (Cité en pages 6, 7, 12, 67, 75 et 80.)
- STEVENS, S. (1956). Calculation of the loudness of complex noise. *The Journal of the Acoustical Society of America*, 28(5) : 807–832. (Cité en page 22.)
- STEVENS, S. (1957). On the psychophysical law. *Psychological review*, 64(3) : 153–181. (Cité en page 78.)
- STEVENS, S. (1972). Perceived level of noise by Mark VII and Decibels (E). *The Journal of the Acoustical Society of America*, 51(2) : 575–601. (Cité en pages 7, 72, 80 et 85.)
- STEVENS, S. (1975). *Psychophysics*. Transaction Publishers. (Cité en page 80.)
- STEVENS, S. et GREENBAUM, H. (1966). Regression effect in psychophysical judgment. *Perception & Psychophysics*, 1(12) : 439–446. (Cité en pages 73, 79 et 80.)
- STEVENS, S. et GUIRAO, M. (1962). Loudness, reciprocity, and partition scales. *The Journal of the Acoustical Society of America*, 34(8) : 1466–1471. (Cité en pages 24 et 79.)
- STEVENS, S. et GUIRAO, M. (1964). Individual loudness functions. *The Journal of the Acoustical Society of America*, 36(11) : 2210–2213. (Cité en pages 62, 80 et 85.)
- SUZUKI, S., SUZUKI, Y., KONO, S., SONE, T., KUMAGAI, M., MIURA, H. et KADO, H. (1989). Equal-loudness level contours for pure tone under free field listening conditions (i) : Some data and considerations on experimental conditions. *Journal of the Acoustical Society of Japan (E)*, 10(6) : 329–338. (Cité en page 8.)
- SUZUKI, Y., ABE, K., OZAWA, K. et SONE, T. (2000). Factors for perceiving sound environments and the effect of visual and verbal information on these factors. In : *Contributions to Psychological Acoustics : Eight Oldenburg Symposium on Psychological Acoustics*, pages 209–232. (Cité en page 25.)

- SUZUKI, Y. et TAKESHIMA, H. (2004). Equal-loudness-level contours for pure tones. *The Journal of the Acoustical Society of America*, 116(2) : 918–933. (Cité en pages 8 et 9.)
- TAKESHIMA, H., SUZUKI, Y., ASHIHARA, K. et FUJIMORI, T. (2002). Equal-loudness contours between 1 kHz and 12.5 kHz for 60 and 80 phons. *Acoust. Sci. & Tech.*, 23(2) : 106–109. (Cité en pages 106 et 111.)
- TAKESHIMA, H., SUZUKI, Y., FUJII, H., KUMAGAI, M., ASHIHARA, K., FUJIMORI, T. et SONE, T. (2001). Equal-loudness contours measured by the randomized maximum likelihood sequential procedure. *Acta Acustica united with Acustica*, 87(3) : 389–399. (Cité en pages 8 et 106.)
- TEGHTSOONIAN, R. et TEGHTSOONIAN, M. (1978). Range and regression effects in magnitude scaling. *Perception & Psychophysics*, 24(4) : 305–314. (Cité en page 79.)
- TOWNSEND, T. H. et GOLDSTEIN, D. P. (1972). Suprathreshold binaural unmasking. *The Journal of the Acoustical Society of America*, 51(2B) : 621–624. (Cité en page 15.)
- VIEMEISTER, N. F. et BACON, S. P. (1988). Intensity discrimination, increment detection, and magnitude estimation for 1-kHz tones. *The Journal of the Acoustical Society of America*, 84(1) : 172–178. (Cité en page 6.)
- WARD, L. (1973). Repeated magnitude estimations with a variable standard : Sequential effects and other properties. *Perception & Psychophysics*, 13(2) : 193–200. (Cité en page 66.)
- WATANABE, T. et MØLLER, H. (1990). Hearing thresholds and equal loudness contours in free field at frequencies below 1 kHz. *Journal of Low Frequency Noise Vibration and Active Control*, 9(4) : 135–148. (Cité en page 8.)
- WERSÉNYI, G. (2010). Representations of HRTFs using MATLAB : 2D and 3D plots of accurate dummy-head measurements. In *Proceedings of the 20th International Congress on Acoustics (ICA), Sydney, Australia*. (Cité en page 60.)
- WHILBY, S., FLORENTINE, M., WAGNER, E. et MAROZEAU, J. (2006). Monaural and binaural loudness of 5-and 200-ms tones in normal and impaired hearing. *The Journal of the Acoustical Society of America*, 119(6) : 3931–3939. (Cité en pages 7, 12, 13, 14 et 15.)
- WIGHTMAN, F. L. et KISTLER, D. J. (1992). The dominant role of low-frequency interaural time differences in sound localization. *The Journal of the Acoustical Society of America*, 91(3) : 1648–1661. (Cité en page 102.)
- ZAHORIK, P. et WIGHTMAN, F. L. (2001). Loudness constancy with varying sound source distance. *Nature neuroscience*, 4(1) : 78–83. (Cité en pages 19, 24, 35, 61, 62 et 138.)
- ZHANG, J. et MAO, D. (2010). Dependence of binaural loudness summation on interaural level difference and frequency for pure tones. *Science China Physics, Mechanics and Astronomy*, 53(5) : 834–841. (Cité en pages 15 et 16.)
- ZIMMER, K. (2005). Examining the validity of numerical ratios in loudness fractionation. *Perception & Psychophysics*, 67(4) : 569–579. (Cité en page 73.)

- ZWICKER, E. (1958). Über psychologische und methodische grundlagen der lautheit. *Acta Acustica united with Acustica*, 8(Supplement 1) : 237–258. (Cité en page 26.)
- ZWICKER, E. (1961). Subdivision of the audible frequency range into critical bands (frequenzgruppen). *The Journal of the Acoustical Society of America*, 33(2) : 248. (Cité en pages 10 et 28.)
- ZWICKER, E. et FASTL, H. (1999). Psychoacoustics, facts and models. (Cité en pages 1 et 10.)
- ZWICKER, E., FASTL, H. et DALLMAYR, C. (1984). BASIC-Program for calculating the loudness of sounds from their 1/3-oct band spectra according to ISO 532 B. *Acustica*, 55 : 63–67. (Cité en page 118.)
- ZWICKER, E., FLOTTORP, G. et STEVENS, S. S. (1957). Critical band width in loudness summation. *The Journal of the Acoustical Society of America*, 29(5) : 548–557. (Cité en page 10.)
- ZWICKER, E. et ZWICKER, U. (1991). Dependence of binaural loudness summation on interaural level differences, spectral distribution, and temporal distribution. *The Journal of the Acoustical Society of America*, 89 : 756. (Cité en pages 2, 13, 15, 16, 29, 34 et 61.)
- ZWISLOCKI, J. (1960). Theory of temporal auditory summation. *The Journal of the Acoustical Society of America*, 32(8) : 1046–1060. (Cité en page 11.)
- ZWISLOCKI, J. (1965). Analysis of some auditory characteristics. In : Luce RD, Bush RR, Galanter E (eds), *Handbook of Mathematical Psychology*. New York : Wiley, pages 1–97. (Cité en page 6.)
- ZWISLOCKI, J. et HELLMAN, R. (1960). On the "psychophysical law". *The Journal of the Acoustical Society of America*, 32(7) : 924–924. (Cité en page 6.)

FOLIO ADMINISTRATIF

THÈSE SOUTENUE DEVANT L'INSTITUT NATIONAL DES SCIENCES APPLIQUÉES DE LYON

NOM : VANNIER

DATE de SOUTENANCE : 11 mai 2015

Prénoms : Michaël, Julien, Marcel

TITRE : SONIE DE CHAMPS ACOUSTIQUES STATIONNAIRES EN SITUATION D'ÉCOUTE DICHOTIQUE

NATURE : Doctorat

Numéro d'ordre : 2015-ISAL-0040

Ecole doctorale : Mécanique, Energétique, Génie civil, Acoustique (MEGA)

Spécialité : Acoustique

RESUME : Dans un environnement naturel, le champ acoustique est complexe (plusieurs sources, différentes positions spatiales, acoustique du lieu,...) et l'écoute est dichotique. La thèse s'est ainsi intéressée à la sonie des champs acoustiques stationnaires, impliquant une ou plusieurs sources, artificielles ou réelles, en champ libre ou dans des acoustiques naturelles. De nouveaux éléments ont été apportés dans la compréhension dont l'information contenue dans les signaux reçus aux oreilles de l'auditeur est combinée pour former un unique percept de sonie binaurale dans les situations d'écoute dichotiques (gain de sommation binaural, cas de plusieurs sources, effet de la corrélation interaurale,...). D'une part, une hypothèse pour essayer d'expliquer les différences interindividuelles observées dans les stratégies de sommation binaurales a pu être testée ; la robustesse et la stabilité au cours du temps de ces stratégies individuelles a été mise en avant. D'autre part, trois principaux modèles psychophysiques de sonie binaurale (ANSI-S3.4 (2007), Moore et Glasberg (2007), Sivonen et Ellermeier (2008)) ont été testés sur l'ensemble des données expérimentales (impliquant différents niveaux de réalisme), permettant de préciser la performance et les domaines de validité respectifs de chacun de ces modèles dans des situations d'écoute fortement dichotiques.

MOTS-CLÉS : Perception, Sonie, Binaural, Dichotique

Laboratoire (s) de recherche : Laboratoire Vibrations Acoustique (LVA)

Directeur de thèse : Etienne Parizet

Président de jury : Daniel Pressnitzer

Composition du jury :

Wolfgang Ellermeier	(rapporteur)
Nicolas Grimault	(rapporteur)
Vincent Koehl	
Sabine Meunier	
Etienne Parizet	
Daniel Pressnitzer	

